**Article:**

Guye, S., De Simoni, C. and von Bastian, C.C. (2017) Do individual differences predict change in cognitive training performance? A latent growth curve modeling approach. Journal of Cognitive Enhancement. ISSN 2509-3290

https://doi.org/10.1007/s41465-017-0049-9

**The final publication is available at link.springer.com via**

**http://dx.doi.org/ 10.1007/s41465-017-0049-9**

Do Individual Differences Predict Change in Cognitive Training Performance?

A Latent Growth Curve Modeling Approach

Sabrina Guye

Carla De Simoni

University of Zurich

Claudia C. von Bastian

Bournemouth University

Author Note

Sabrina Guye, University Research Priority Program (URPP) “Dynamics of Healthy Aging”, University of Zurich, Switzerland; Carla De Simoni, Department of Psychology, University of Zurich, Switzerland; Claudia C. von Bastian, Department of Psychology, Bournemouth University, United Kingdom.

Claudia C. von Bastian is now at the Department of Psychology, University of Sheffield, United Kingdom.

Data reported in this work has been collected with the support of grants awarded to the first and second author from the Suzanne and Hans Biäsch Foundation for Applied Psychology

(Ref 2014/32; 2016/08). The first author was further supported by the Forschungskredit of the University of Zurich (FK-16-062), and the second author by the Swiss National Science Foundation (No. 100014_146074). Moreover, both authors were supported by the URPP "Dynamics of Healthy Aging" of the University of Zurich.

During the work on her dissertation, Sabrina Guye was a pre-doctoral fellow of the International Max Planck Research School on the Life Course (LIFE; participating institutions: MPI for Human Development, Humboldt-Universität zu Berlin, Freie Universität Berlin, University of Michigan, University of Virginia, University of Zurich).

Correspondence concerning this article should be addressed to Sabrina Guye, University of Zurich, University Research Priority Program "Dynamics of Healthy Aging", Andreasstrasse 15, CH-8050 Zürich, Switzerland (e-mail: sabrina.guye@uzh.ch)

Abstract

Cognitive training interventions have become increasingly popular as a potential means to cost-efficiently stabilize or enhance cognitive functioning across the lifespan. Large training improvements have been consistently reported on the group level, with, however, large differences on the individual level. Identifying the factors contributing to these individual differences could allow for developing individually-tailored interventions to boost training gains. In this study, we therefore examined a range of individual differences variables that had been discussed in the literature to potentially predict training performance. To estimate and predict individual differences in the training trajectories, we applied Latent Growth Curve models to existing data from three working memory training interventions with younger and older adults. However, we found that individual differences in demographic variables, real-world cognition, motivation, cognition-related beliefs, personality, leisure activities, and computer literacy and training experience were largely unrelated to change in training performance. Solely baseline cognitive performance was substantially related to change in training performance and particularly so in young adults, with individuals with higher baseline performance showing the largest gains. Thus, our results conform to magnification accounts of cognitive change.

Keywords: working memory training, individual differences, latent growth curve modeling

# Do Individual Differences Predict Change in Cognitive Performance? A Latent Growth Curve Modeling Approach

Over the past decade, there has been an exploding interest in computer-based commercial "brain training" programs and in scientific evidence relating to the effectiveness of such interventions, triggered by promising results of working memory (WM) training gains generalizing to previously untrained cognitive abilities such as intelligence in both younger (e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, 2008) and older adults (e.g., Borella, Carretti, Riboldi, & De Beni, 2010). Although the idea of improving general cognitive functioning within a few weeks is enticing, there is also accumulating evidence against a generalized effect of WM training (e.g., Clark, Lawlor-Savage, & Goghari, 2017; De Simoni & von Bastian, 2017; Guye & von Bastian, 2017; Sprenger et al., 2013). Even on the meta-analytic level, evidence is mixed regarding the effectiveness of cognitive training in both younger and older adults (e.g., Au et al., 2015; Dougherty, Hamovitz, & Tidwell, 2016; Karbach & Verhaeghen, 2014; Kelly et al., 2014; Lampit, Hallock, & Valenzuela, 2014; Melby-Lervåg & Hulme, 2013; Melby-Lervåg, Redick, & Hulme, 2016; Schwaighofer, Fischer, & Bühner, 2015; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017). Aside from design and methodological choices potentially explaining the diverging findings (e.g., Noack, Lövdén, Schmiedek, & Lindenberger, 2009; Shipstead, Redick, & Engle, 2012), many authors increasingly articulated the potentially important influence of individual differences on cognitive training trajectories and outcomes (e.g., Buitenweg, Murre, & Ridderinkhof, 2012; Guye, Röcke, Mérillat, von Bastian, & Martin, 2016; Könen & Karbach, 2015; Shah, Buschkuehl, Jaeggi, & Jonides, 2012; von Bastian & Oberauer, 2014)

Individual differences in cognitive functioning (e.g., Ackerman & Lohman, 2006) and learning potential (e.g., Stern, 2017) accentuate with increasing age (e.g., Rabbitt, Diggle, Holland, & McInnes, 2004), and have been shown to be related to personality (e.g., Graham & Lachman, 2012), cognition-related beliefs such as need for cognition (NFC; e.g., Fleischhauer et al., 2010; Hill et al., 2013), and everyday life activities (e.g., Jopp & Hertzog, 2007). Investigating which of these individual differences potentially predict cognitive training outcomes may not only explain inconsistencies concerning the effectiveness of cognitive training, but also identify possible subgroups of individuals that are more or less responsive to cognitive training, thereby constituting the conceptual groundwork for developing individually-tailored interventions to boost training effectiveness.

**Predictors of Cognitive Training Outcomes**

As yet, only few studies have examined how individual differences are associated with cognitive training outcomes (see Katz, Jones, Shah, Buschkuehl, & Jaeggi, 2016 for an overview), with most existing studies relating training outcomes to demographic variables (e.g., age), baseline cognitive performance, motivation, cognition-related beliefs (e.g., theories of intelligence; TIS) and personality traits (e.g., neuroticism and conscientiousness).

So far, the effect of age on training outcomes has received the most attention. Age-comparative studies mostly reported larger training effects in younger than in older adults (e.g., Brehmer, Westerberg, & Bäckman, 2012; Bürki, Ludwig, Chicherio, & de Ribaupierre, 2014; Schmiedek, Lövdén, & Lindenberger, 2010; von Bastian, Langer, Jäncke, & Oberauer, 2013), and in young-old adults compared to old-old adults (e.g., Borella et al., 2014; Zinke et al., 2014). These results are in line with the notion of a magnification effect (also known as amplification or Matthew effect; Kliegl, Smith, & Baltes, 1990; Lövdén, Brehmer, Li, & Lindenberger, 2012; Verhaeghen & Marcoen, 1996), suggesting that younger individuals benefit more from cognitive training, as they have the additional cognitive resources available required for successfully completing the training tasks. However, other studies found that children and older adults benefited more from training than young adults (e.g., Bherer et al., 2008; Karbach & Kray, 2009). Such compensation effects have been argued to emerge as participants with lower initial cognitive status have more room for improvement (see Titz & Karbach, 2014 for a review). These diverging findings are reflected by recent meta-analyses, with some reporting evidence for age being a moderator of training outcomes (e.g., Melby-Lervåg & Hulme, 2013) and others not (e.g., Karbach & Verhaeghen, 2014; Schwaighofer et al., 2015). A closely related, yet potentially distinct factor possibly contributing to these mixed findings is general cognitive functioning (von Bastian & Oberauer, 2014). Only few studies have directly assessed the effect of baseline cognitive performance on training outcomes though, with some evidence suggesting that initially low-performing individuals benefit more from training (e.g., Jaeggi et al., 2008; Zinke et al., 2014), but others reported opposite effects (e.g., Bürki et al., 2014).

Although motivation is arguably one of the most plausible factors possibly influencing cognitive training outcomes, its association with training performance has not yet been comprehensively examined. One exception is a study by Brose, Schmiedek, Lövdén, and Lindenberger (2012), who reported a positive association between daily motivation and daily cognitive performance on a 3-back task, indicating that on days on which task-related motivation was lower than on average, daily cognitive performance was also reduced. Some studies have investigated the effect of related concepts, including cognition-related beliefs such as individuals’ beliefs about the malleability of intelligence (TIS; Dweck, 2000). For instance, Jaeggi, Buschkuehl, Shah, and Jonides (2014) found that, irrespective of training intervention (control or experimental intervention), the group of individuals indicating high beliefs in the malleability of intelligence (a “growth mindset”) showed larger transfer effects than the group of individuals who believed that intelligence cannot be changed (but see Thompson et al., 2013). Due to the fact that the groups were determined by median split, these results should, however, be interpreted with caution, as median split and extreme group analyses can potentially inflate the effect sizes and consequently overestimate the importance of a given effect (Moreau, Kirk, & Waldie, 2016; Unsworth et al., 2015). Indeed, other studies have not found an association of cognition-related beliefs with training outcomes (Minear et al., 2016; Sprenger et al., 2013).

Finally, there is some evidence for personality traits being related to training outcomes. It has been reported that conscientiousness is positively related to training performance, but negatively to far transfer effects (Studer-Luethi, Jaeggi, Buschkuehl, & Perrig, 2012). Further, neuroticism has been found to be negatively associated with mean training performance (but not training gain; Studer-Luethi, Bauer, & Perrig, 2016; Studer-Luethi et al., 2012) and transfer effects (Studer-Luethi et al., 2012; 2016; see also Urbánek & Marček, 2015 for similar results using the Personality System Interaction personality factors), except when training task complexity is low (Studer-Luethi et al., 2012).

In sum, there is some tentative evidence that individual differences may predict training performance and transfer effects. Studies attempting to estimate the role of individual differences based on sufficiently large training samples and continuous predictors are, however, scarce. Further, some individual differences have been entirely neglected, including cognitive performance in real-world context (e.g., education), training-related leisure activities (e.g., gaming), and computer literacy or previous training experience.

## The Present Study

The goal of this study was to enhance the understanding of who benefits from cognitive training and who does not. Using Latent Growth Curve (LGC) modeling, we therefore examined

(1) the individual cognitive training trajectories, (2) the association of baseline cognitive performance with change in training performance, and (3) which individual differences predicted change in training performance.

We reanalyzed three data sets obtained from two randomized-controlled, double-blind WM training studies investigating two WM interventions in younger (De Simoni & von Bastian, 2017) and one in older adults (Guye & von Bastian, 2017). Observed improvements in the trained tasks were substantial in size and in line with numerous studies consistently reporting training effects across a wide variety of training regimes and trained abilities (e.g., Karbach & Verhaeghen, 2014). The two training studies were similar regarding the included questionnaires assessing individual differences potentially predicting training performance, and the training regimen itself (i.e., trained tasks, training duration, frequency, adaptive task difficulty, and nature of the control group). In the first study (De Simoni & von Bastian, 2017), younger adults received either of two single-paradigm WM training interventions (i.e., memory updating and binding training). In the second study (Guye & von Bastian, 2017), older adults received a mixed-paradigm WM training intervention, consisting of a memory updating, a binding, and a complex span task. All three interventions were adaptive, with the level of difficulty increasing depending on individuals' performance.

To estimate the training trajectories, we fitted LGC models to the data recorded during training. LGC modeling uses structural equation modeling (SEM) to estimate interindividual differences in intraindividual change over time. LGC modeling is highly flexible as it can handle a variety of methodological issues typically occurring in training research such as partially missing data, non-normally distributed data, or non-linear change trajectories (Curran, Obeidat, & Losardo, 2010). Further, LGC modeling has the advantage to account for measurement error and to provide separate latent estimates for baseline cognitive performance (i.e., the intercept) and change in training performance (i.e., the slope). The distinction between the two latent factors allows for estimating how baseline cognitive performance is related to change in performance, with a positive relationship reflecting magnification, and a negative relationship reflecting compensation effects. Further, to investigate how the individual differences variables are associated with the intercept and the slope, we extended the LGC models by predicting the variance in baseline cognitive performance and, more importantly, change in training performance by (1) demographic variables, (2) real-world cognition, (3) motivation, (4) cognition-related beliefs, (5) personality, (6) leisure activities, and (7) computer literacy and training experience.

Statistical evidence for the predictive value of baseline cognitive performance and each of the individual differences variables was evaluated using Bayes factors (BF). The BF is a statistical index ranging from 0 to infinity and quantifies the strength of evidence for one hypothesis (usually the alternative hypothesis $H_1$, postulating the presence of an association) compared to another hypothesis (usually the null hypothesis $H_0$, postulating the absence of an association). Hence, BFs allow for evaluating the strength of evidence not only for the presence of an association, but explicitly also for the absence of a proposed association. Accordingly, using BFs has become increasingly popular in the area of cognitive enhancement (e.g., Antón et al., 2014; Clark, Lawlor-Savage, & Goghari, 2017; De Simoni & von Bastian, 2017; Guye & von Bastian, 2017; Kirk, Fiala, Scott-Brown, & Kempe, 2014; Paap, Johnson, & Sawi, 2014; Sprenger et al., 2013; von Bastian, Guye, & De Simoni, 2017; von Bastian & Oberauer, 2013).

Based on previous findings, we expected positive associations of motivation (Brose et al., 2012), a growth mindset (Jaeggi et al., 2014), and conscientiousness (Studer-Luethi et al., 2012) with change in training performance. Regarding neuroticism, our expectations were less specific, given that previous literature reported evidence for a negative association of neuroticism with mean training performance and transfer effects, but not with training gains (e.g., Studer-Luethi et al., 2012; 2016). Based on the results by Bürki et al. (2014), methodologically the most similar study to our own, we expected a negative association of age and a positive association of baseline cognitive performance with change in cognitive performance, which would support the magnification hypothesis. For all the other individual differences variables, the analyses were exploratory.

## Method

Detailed methods regarding the training interventions have been reported previously (De Simoni & von Bastian, 2017; Guye & von Bastian, 2017). In the following, we summarize the key characteristics of each study's methodology with a focus on the individual differences measures.

### Participants

The final sample sizes ranged from 58 to 68 (see Table 1 for detailed sample description). The Young-Updating and Young-Binding samples were drawn from a study of healthy younger participants aged between 18 – 36 years, and the Old-Mixed sample was drawn from a study of healthy older participants aged between 65 – 80 years. Younger participants were recruited through the participant pool of the Department of Psychology of the University of Zurich, postings at the university campus, and short study presentations during lectures. Older participants were recruited through the participant pool of the University Research Priority Program "Dynamics of Healthy Aging", lectures at the Senior Citizens' University of Zurich, flyers, online announcements, and word-of-mouth. All participants were fluent or native German speakers and had a computer with Internet connection at home. Written informed consent was obtained from all participants. Both studies were approved by the ethics committee of the Department of Psychology of the University of Zurich. After study completion, younger participants received either CHF 120 (approx. USD 120) or CHF 20 (USD 20) plus 10 course credits, moreover, they could earn a bonus up to a maximum of 50 CHF (USD 50), depending on the level of difficulty that they reached during training. Older participants received CHF 150 (approx. USD 150).

Younger participants reported no current psychiatric or neurological disorders, psychotropic drug use, or color blindness. Older participants also reported no current psychiatric or neurological disorders, psychotropic drug use, and no significant motor, hearing or vision impairments. Further, they were screened for color blindness (Ishihara, 1917), subclinical depression (GDS; Sheikh & Yesavage, 1986: cut-off criterion = 4), and cognitive impairment (MMSE; Folstein, Folstein, & McHugh, 1975: cut-off criterion = 26).

Table 1

Demographics of Study Participants

| | Sample | | |
|---|---|---|---|
| Demographics | Young-Updating | Young-Binding | Old-Mixed |
| Sample size (n) | 58 | 64 | 68 |
| Intervention | Memory updating | Binding | Mixed-paradigm |
| Age | 22.57 (2.99) | 24.77 (4.03) | 70.40 (3.72) |
| Gender (f/m) | 39/19 | 45/19 | 30/38 |
| Education [a] | 5 (0.00) | 5 (0.00) | 5 (1.48) |
| MMSE score | - | - | 29.21 (0.76) |
| GDS score | - | - | 0.65 (1.02) |

Note. Values are means and standard deviations in parentheses (median and median absolute deviation in parentheses for education).

[a] The scale for education ranged from 0 (no formal education) to 7 (doctorate).

## Studies and Material

**Cognitive training interventions.** Training procedures were identical for the three samples if not mentioned otherwise. Tatool was used to deliver the self-administered training interventions at home and to monitor participants' training compliance (von Bastian, Locher, & Ruflin, 2013). The default adaptive score and level handler implemented in Tatool was used to adjust task difficulty to participants' performance throughout the training phase. Both the set size (i.e., number of memoranda) and the response time limit varied depending on the level of task difficulty (see below). Younger participants completed 20 sessions of WM training (30-45 minutes per session) within five weeks. Each training session consisted of 12 trials per task in the Young-Updating sample and 24 trials per task in the Young-Binding sample. Interventions comprised verbal, spatial, visual, and numerical memory updating tasks (Young-Updating sample) and verbal, spatial, visual, and numerical binding tasks (Young-Binding sample). Both younger samples trained each task for a maximum of 11.25 min per session. Older participants completed 25 sessions of WM training (30-45 minutes per session) within five weeks, with the intervention consisting of a complex span, a binding, and a memory updating task each of which contained visuo-spatial memoranda. Each task was trained for a maximum of 15 min per session, with each session consisting of 15 trials per task. Set size achieved at the end of each session and task was used as the dependent variable. Table 2 lists an overview of the training tasks.

Table 2

Working Memory Training Tasks of the Training Interventions

| Task(-version) | Description |
|---|---|
| | Memory updating training |
| Arrows | Memorize a set of arrows and update by rotating them for 45 degrees clockwise or counterclockwise. |
| Letters | Memorize a set of letters and update by mentally shifting them up to three positions forward or backward in the alphabet. |
| Locations | Memorize the locations of a set of circles in a grid and update by mentally shifting them to an adjacent cell as indicated by an arrow. |
| Digits | Memorize a set of digits and update by applying simple arithmetic operations to them. |
| | Binding training |
| Fractal-location | Memorize a series of associations between fractals and their location in a row of boxes on the grid. |
| Noun-verb | Memorize a series of associations between nouns and verbs. |
| Color-location | Memorize a series of associations between colored circles and their locations in a 4 x 4 grid. |
| Symbol-digit | Memorize a series of associations between mathematical symbols and digits. |
| | Mixed-paradigm training |
| Memory updating | Memorize the locations of a set of circles in a 4 x 4 grid and update by mentally shifting them to an adjacent cell. |
| Binding | Memorize a series of associations between colored triangles and their locations in a 4 x 4 grid. |
| Complex span | Memorize a series of positions of squares in a 5 x 5 grid interleaved by a distractor task. |

Note. Detailed description of the tasks can be found in the original publications (De Simoni & von Bastian, 2017; Guye & von Bastian, 2017).

Updating training. The Young-Updating sample practiced four memory updating tasks (adapted from Lewandowsky, Oberauer, Yang, & Ecker, 2010). In these tasks, participants had to memorize a set of stimuli presented simultaneously for 500 ms per item. In the subsequent updating phase, participants had to transform individual memoranda (e.g., mentally rotate previously memorized arrows or applying a simple arithmetic operation to a number), enter the result of the transformation, and remember that result of the transformation. In half of the trials, a cue presented for 500 ms indicated which of the memorandum had to be updated. After nine updating steps, participants had to recall the most recent result of each stimulus. Task difficulty was adjusted to individual performance by increasing the set size (i.e., number of simultaneously presented memoranda) and reducing the time limit to respond to the updating prompts.

Binding training. The Young-Binding sample practiced four binding tasks (adapted from Wilhelm, Hildebrandt, & Oberauer, 2013). In these tasks, participants had to remember associations between elements (e.g., noun and verbs or objects and their locations in a grid) presented sequentially for 900 ms (noun-verb and symbol-digit) or 1800 ms (fractal-location and color-location) each. After memorization, each association was probed in random order with one of the elements given as cue. Half of the probes were positive (i.e., exact matches), whereas negative probes could be distractors (i.e., probes not presented in the current trial; 25 % of probes) or intrusions (i.e., probes that were presented in the current trial, but associated with a different element; 25 % of probes). Task difficulty was adjusted to individual performance by increasing the set size (i.e., number of sequentially presented pairs) and reducing the time limit to respond to the probes.

Mixed-paradigm training. Mixed-paradigm training consisted of a memory updating task (adapted from De Simoni & von Bastian, 2017; Schmiedek, Lövdén, & Lindenberger, 2014), a binding task (Oberauer, 2005), and a figural-spatial complex span task (von Bastian & Eschen, 2016).

The memory updating task was identical to the locations task practiced by the Young-Updating sample. Participants first had to memorize the locations of colored circles presented simultaneously in a 4 x 4 grid for 500 ms per item. After the presentation of the circles, an arrow was presented alongside one of the circles centrally on the screen for 500 ms. The circle had to be mentally shifted up, down, left, or right to the adjacent cell as indicated by the arrow. Participants indicated the new position of the circle by mouse click in the blank grid. As in the Young-Updating Sample updating training, trials comprised nine updating steps, with half of the trials using a cue presented for 500 ms to indicate which of the circles had to be updated.

The binding task was similar to the ones practiced by the Young-Binding sample. Participants had to memorize a series of locations of colored triangles in a 4 x 4 grid. Each item was presented for 900 ms followed by a 100 ms inter-stimulus interval. During recognition, each association was probed by presenting a triangle in a location in the grid, and participants had to decide whether it matched the triangle that was previously presented at that position. Across all trials, 50 % of the probes were matches, 25 % were distractors, and 25 % were intrusions.

For the complex span task, participants had to memorize a series of red in a 5 x 5 grid, each presented for 1000 ms. Each trial of the series was interleaved by a distractor task, in which participants had to decide whether the long side of a L-shaped figure within the grid was oriented vertically or horizontally. Response time during the distractor task was limited to 3000 ms. During recall, participants had unlimited time to indicate the grid positions in correct serial order by mouse-click.

In all three tasks, difficulty was adjusted by increasing the set size and reducing the response time limit. For the complex span task, time to respond to the distractor task was limited, and for the binding and memory updating tasks time to respond during the retrieval phase was reduced.

Adaptive task difficulty. All participants started training on the same level of task difficulty. To maximize the time participants were exposed to challenging task demands, we ensured that participants quickly reached their individual baseline cognitive performance limit by implementing a fast evaluating adaptive algorithm during the first training session. Participants' performance was evaluated after every 10 % of trials in the younger samples, and every 7 % of trials in the older sample (corresponding to one trial in the Young-Updating sample and the Old-Mixed sample, and two trials in the Young-Binding sample). If participants reached a performance criterion (i.e., accuracy above 85 % in the younger samples, 80 % in the older sample), task difficulty was raised by reducing the response time limit (by 500 ms in the younger samples and 300 ms in the older sample) for four subsequent level-ups, or by increasing the set size by one additional memorandum every fifth level-up (which also reset the response time limit to the starting value). After the first session, performance was evaluated after every 40% of trials (corresponding to five trials in the Young-Updating sample, ten trials in the Young-Binding sample and six trials in the Old-Mixed sample). The first training session started with a set size of two and a response time limit of 3500 ms per response for the younger samples, and 5000 ms per response in the older sample. The maximum set size was set to eight in the Young-Updating and the Old-Mixed samples and seven in the Young-Binding sample.

**Assessment of individual differences variables.** Individual differences variables were assessed prior to training, except for motivation, which was assessed at the end of the respective training sessions (see below). Participants completed most computer-based questionnaires at home. In addition, older adults completed the following questionnaires during an individual in-lab assessment at the University of Zurich: a demographic questionnaire, a computer- and Internet questionnaire, and an adapted German, multiple-choice version of the Everyday Performance Test (EPT; Willis & Marsiske, 1993). Mean rating was used as the dependent variable for the questionnaire measures.

**Demographics.** Age and gender were assessed with a demographic questionnaire.

**Real-world cognition.** Education level was assessed on a scale ranging from 0 to 7 (0 = no formal education, 7 = doctorate). As younger adults were only included in the study if they obtained at least a higher education entrance qualification (corresponding to education level 4), variance in this measure was limited. Thus, we refrained from using education level as a predictor in younger adults. Older adults additionally completed the Cognitive Failure Questionnaire (CFQ; Broadbent, Cooper, FitzGerald, & Parkes, 1982), assessing self-reported failures in perception, memory, and motor function. Items such as “Do you find you forget people’s names?” were rated on a 5-point scale (0 = never, 4 = very often). Further, we assessed older adults’ everyday problem solving abilities using an adapted multiple-choice version of the EPT. The EPT is an objective assessment of everyday competence to perform complex tasks of daily living. Participants were presented with 15 everyday tasks (e.g., a recipe for twelve biscuits) and asked to solve two problems associated with each stimulus (e.g., calculate the amount of flour to bake half of the biscuits) by choosing one of four answers. EPT score represents the number of correctly solved items within 45 minutes.

**Motivation.** In the younger samples, participants’ training motivation was assessed at the beginning of and mid-way through training (sessions 1 and 10) using an adapted version of the

Questionnaire on Current Motivation (Rheinberg, Vollmeyer, & Bruns, 2001). On a 7-point scale (1 = disagree, 7 = agree) they had to rate items such as "I am fully determined to give my best during training". In addition, the younger participants completed an adapted version of the Intrinsic Motivation Inventory (IMI; Deci & Ryan, 2016) at the end of the last training session, rating items such as "Today's training session was fun to do" on a 7-point scale (1 = does not apply at all, 7 = does apply very well). In the older sample, participants' training motivation was assessed at the beginning of and mid-way through training (sessions 2 and 14) using an adapted version the IMI (Deci and Ryan, 2016). Because the motivation measures were highly correlated in the younger (all rs $\geq$ .48, all ps < .001) and older samples (r = .76, p < .001) across time points, we computed one single motivation composite score by averaging the z-transformed scores.

**Cognition-related beliefs.** Beliefs were measured using four different constructs. First, we assessed participants' passion and perseverance for long-term goals using the 12-item Grit scale (Duckworth, Peterson, Matthews, & Kelly, 2007). Items such as "I finish whatever I begin" were rated on a 5-point scale (1 = not like me at all, 5 = very much like me). Second, we assessed the degree to which participants enjoy effortful cognitive activities using the 16-item[1] NFC scale (Cacioppo & Petty, 1982). Items (e.g., "I really enjoy a task that involves coming up with new solutions to problems") were rated on a 7-point scale (1 = strongly disagree, 7 = strongly agree). Third, participants' implicit beliefs about the malleability of intelligence was assessed using the TIS (Dweck, 2000). Items such as "No matter who you are, you can significantly change your intelligence level" were rated on a 6-point scale (1 = strongly disagree, 6 = strongly agree). Higher levels indicate an incremental view (a "growth mindset", i.e., viewing intelligence as a

[1] In the older sample, the 33-item version was administered. To match the younger samples, we only included the 16 items from the short version in the present analyses.

malleable, changeable construct). Finally, to assess participants' sense of perceived self-efficacy, we administered the General Self-Efficacy scale (GSE; Schwarzer & Jerusalem, 1995). Participants rated the items (e.g., "I can always manage to solve difficult problems if I try hard enough") on a 4-point scale (1 = not at all true, 4 = exactly true). Younger adults additionally completed an adapted version of the Self-Efficacy to Regulate Exercise scale (EXSE; Bandura, 2006). Participants rated the items (e.g., "How certain are you that you can get yourself to perform your training routine regularly when you have other time commitments") on a visual analogue scale ranging from 1 to 100.[2]

**Personality.** Personality traits were assessed using the 60-item NEO Five Factor Inventory (Costa & McCrae, 1992), including subscales for neuroticism, agreeableness, openness, conscientiousness, and extraversion. All items were rated on a 5-point scale (0 = strongly disagree, 4 = strongly agree).

**Leisure activities.** Leisure activities were assessed using an adapted version of the Adult Leisure Activity Questionnaire (Jopp & Hertzog, 2010). Items were grouped into 11 activity categories (i.e., physical, developmental, and experiential activities, activities with close social partners, group-centered public activity, religious activities, crafts, game playing, TV watching, travel, and technology use), across which participants indicated how often they partook in these activities on a 6-point scale (1 = never, 6 = daily).

**Computer literacy and training experience.** Older participants completed a questionnaire regarding their computer and Internet experience. Participants were asked "How confident do you feel using the computer?" and responded on a 7-point scale (1 = not confident at all, 7 = very confident). Further, participants were asked if they had any previous cognitive training

[2] As the two measures for self-efficacy were not correlated ($r = 0.03$, $p = .715$), we analyzed both measures separately rather than computing a composite score.

experience (i.e., through commercially available training programs and/or through participating in other studies).

**Data Analysis**

We fitted LGC models to the training data (1) to estimate the individual trajectories of performance change over time and (2) to investigate the effect of baseline cognitive performance on change in training performance, and (3) to identify possible individual differences that predict change in training performance. Ideally, all training sessions would have been included individually in the models (see also Bürki et al., 2014). However, due to the relatively small sample sizes and to increase the signal-to-noise ratio, we reduced the data to five training blocks for each sample by averaging across four sessions in the younger adults (i.e., sessions 1-4, 5-9, 10-14, 15-20) and five sessions in the older adults (i.e., sessions 1-5, 6-10, 11-15, 16-20, 21-25). Further, as we were interested in estimating and predicting general rather than task-specific WM training performance, we used an average of the set size achieved at the end of each session across the four binding or memory updating tasks in the younger adults, and across the three training tasks in the older adults as dependent measure.

By modeling two latent variables, the intercept and the slope, LGC modeling allows for parsimoniously describing both linear and non-linear longitudinal trajectories within the SEM framework by accounting for error variance in the manifest variables. Whereas the value in the dependent variable at the beginning of training ($\mu_i$ = baseline cognitive performance) is represented by the intercept, the rate of change in the dependent variable ($\mu_s$ = increase/decrease in cognitive training performance) is expressed by the slope. Both latent factors are defined by a set of manifest variables (i.e., the training blocks). The model further allows for individual variation in the intercept ($\sigma^2_i$ = variance in baseline cognitive performance) and the slope ($\sigma^2_s$ =

variance in change of training performance), and this variance can in turn be predicted by additional variables (i.e., individual differences). The covariance between the intercept and the slope ($\sigma_{i,s}$) indicates the degree to which baseline performance and change of training performance are correlated, with a positive covariation supporting a magnification effect, and a negative covariation supporting a compensation effect. Finally, the model includes error covariances ($\sigma_{\varepsilon,\varepsilon}$) accounting for correlated error terms ($\varepsilon_{1\text{-}5}$) between the adjacent training blocks. Error variances ($\sigma^2\varepsilon_{1\text{-}5}$) were constrained to be equal across the five error terms.

Model fit was evaluated using the chi-square statistic ($\chi^2$), the standardized root-mean-square residual (SRMR), and the comparative fit index (CFI). Conventionally, good fit is indicated by values between 0 and 2df for the $\chi^2$, by values smaller than 0.08 for the SRMR and greater than 0.95 for the CFI (Hu & Bentler, 1999; Schermelleh-Engel, Moosbrugger, & Müller, 2003). Although the root-mean-square error of approximation (RMSEA) is a popular measure of goodness-of-fit, we do not report it following the recent suggestion of Kenny, Kaniskan, and McCoach (2015). Using Monte Carlo simulations, they showed that the RMSEA tends to over-reject properly specified models with small degrees of freedom, which is the case for all our baseline models (dfs = 7).

All analyses were conducted in R (version 3.2.3; R Core Team, 2015) using the “lavaan” package (version 0.5.23; Rosseel, 2012). Figures depicting training performance were conducted using the “longCatEDA” package (version 0.31; Tueller, Van Dorn, & Bobashev, 2016). The package depicts categorical longitudinal data (in our case the dependent variable set size) by using shades of color instead of vertical position to indicate changes on categorical variables over time.

## Results

Data and analyses scripts are available on the Open Science Framework (https://osf.io/qgkp2). First, to test whether participants training performance increased over the course of the intervention and whether this increase follows a linear or non-linear pattern, we ran three baseline models for each sample (i.e., a no-growth, a linear growth, and a non-linear growth model). We selected the best fitting model using nested model comparisons. Second, we investigated whether baseline cognitive performance is associated with change in training performance and, if so, in which direction. Third, to examine how individual differences are associated with change in training performance, we included the individual differences variables to predict cognitive training trajectories.

To avoid potential issues caused by multicollinearity of predictors, we ran separate models for (1) demographic variables, (2) real-world cognition, (3) motivation, (4) cognition-related beliefs, (5) personality, (6) leisure activities, and (7) computer literacy and training experience. To estimate multicollinearity within the predictor categories, we assessed the Variance Inflation Factor (VIF) in both younger and older samples. The VIFs indicated no signs of multicollinearity, with the highest VIF = 2.18 (for correlation coefficients of the individual differences see Tables S1 and S2 in the supplemental materials). For each of these seven models, all measures were included simultaneously and regressed on the latent intercept and slope concurrently, although the primary interest lies on the prediction of change in training performance (i.e., the slope). Ordinal and metric predictors were z-transformed prior to data analysis.

**Missing Data**

For data analysis, data were included for all participants who performed above chance level during at least 75 % of training sessions (i.e., ≥ 15 sessions for the younger samples, and ≥ 19 sessions for the older sample). We did not include data from three older participants because they (contrary to the instructions) concurrently trained on two computers on two different levels of difficulty. One older participant had to re-install the training software after six training sessions due to technical issues and we used the following 19 sessions for data analyses.

All participants from the Young-Updating sample completed 20 training sessions. However, due to a programming error, the feedback presented during training was incorrect for two participants for the first 2 and 4 sessions, respectively. Consequently, we treated the data from those sessions as missing. In the Young-Binding sample, most participants completed 20 sessions ($M = 19.83$, $SD = 0.70$, range = 15-20). However, four participants did not complete one training session, one participant did not complete two training sessions, and one participant restarted training after 15 sessions. Therefore, we also treated those sessions as missing. Also, most older participants completed 25 sessions ($M = 24.85$, $SD = 0.98$, range = 19-28), except for three participants who completed less due to scheduling problems (i.e., 21, 23, and 24 sessions) and the one person who re-installed the training software (i.e., 19 sessions). If participants completed more than 25 training sessions, these additional sessions were omitted from data analysis.

As we only had missing data for continuous variables but not for categorical or ordinal variables (e.g., gender or education), missing data were handled using Full Information Maximum Likelihood (FIML) estimation, thereby using all available information for estimating the model (see also Grimm, Ram, & Estabrook, 2017).

**Bayes Factors**

We computed BFs for the effect of each predictor on the slope or intercept, allowing for quantifying the evidence for both the alternative hypothesis (i.e., predictor is associated with slope or intercept) and the null hypothesis (i.e., predictor is not associated with slope or intercept). Further, we computed BFs for the variances of the intercept and the slope, as well as for the covariance between the intercept and the slope. BFs were approximated based on the Bayesian Information Criterion (BIC), which evaluates model fit based on the log-likelihood taking the degrees of freedom into account, with a lower BIC reflecting a better model fit. The BF is computed using the difference in BICs when comparing the model freely estimating the predictor of interest and the model in which the predictor of interest is fixed to zero (Wagenmakers, 2007):

$$BF_{H1} = \exp(0.5*(BIC_2 - BIC_1),$$

with $BIC_1$ being the BIC for the alternative model freely estimating the predictor of interest, and $BIC_2$ being the BIC for the identical model with the predictor of interest fixed to zero (i.e., the null model). BFs range from 0 to infinity, with higher values indicating stronger evidence for the alternative model. BFs are evaluated according to an adapted version of Wetzels and Wagenmakers (2012) to facilitate verbal interpretation (see Table 3). For example, a BF of 3 indicates that the data is three times more likely to occur under the alternative hypothesis. BFs favoring the null model (i.e., BFs < 1) are expressed as 1/BF.

Table 3

Verbal Labels to Guide Interpretation of Bayes Factors

| Bayes factor | Interpretation |
|---|---|
| > 100 | Decisive |
| 30-100 | Very strong |
| 10-30 | Strong |
| 3-10 | Substantial |
| 1-3 | Ambiguous |
| 1 | No evidence |

Note. Adapted from Wetzels and Wagenmakers (2012).

**Specification of the Baseline Model**

To identify the best fitting baseline model, we conducted several nested model comparisons for each sample and assessed whether there was a significant improvement of the relative fit (see Table 4). We compared three models: a no growth curve model assuming no change in cognitive performance (Model 1), a linear model assuming linear change in cognitive performance (Model 2), and a non-linear model assuming non-linear change in cognitive performance (Model 3). Model 3 was modeled according to Kline (2016) by fixing the first two coefficients of the slope factor to constants (0, 1) and freeing the remaining coefficients for the slope factor. This specification allows for estimating an empirical curvilinear trend that optimally fits the data. For all samples, Model 3 fitted the data significantly better than Models 1 and 2.

Table 4

Nested Model Comparisons and Fit Indices for Baseline Latent Growth Curve Models

| | $\chi^2$ | df | SRMR | CFI | Model comparison | $\Delta\chi^2$ | $\Delta$df |
|---|---|---|---|---|---|---|---|
| Young-Updating | | | | | | | |
| Model 1 | **435.47** | 13 | 1.15 | .22 | - | - | - |
| Model 2 | **52.56** | 10 | 0.08 | .92 | 1 vs. 2 | **382.91** | 3 |
| Model 3 | 4.04 | 7 | 0.02 | 1.00 | 2 vs. 3 | **48.52** | 3 |
| Young-Binding | | | | | | | |
| Model 1 | **534.73** | 13 | 1.79 | .12 | - | - | - |
| Model 2 | **142.11** | 10 | 0.16 | .78 | 1 vs. 2 | **392.62** | 3 |
| Model 3 | **23.22** | 7 | 0.04 | .97 | 2 vs. 3 | **118.89** | 3 |
| Old-Mixed | | | | | | | |
| Model 1 | **413.89** | 13 | 0.82 | .23 | - | - | - |
| Model 2 | **32.88** | 10 | 0.08 | .96 | 1 vs. 2 | **381.01** | 3 |
| Model 3 | 11.83 | 7 | 0.05 | .99 | 2 vs. 3 | **21.06** | 3 |

Note. Bold values represent significant $\chi^2$ statistics ($p < .05$)

**Latent Analysis of Training Performance**

Results for the baseline models are summarized in Figure 1. Training performance for each training task is visualized in Figure 2 for younger adults, and Figure 3 for older adults. Training performance across tasks for the three samples is visualized in Figure 4.

The non-linear baseline LGC model fitted the data from the Young-Updating sample well, $\chi^2(7) = 4.04$, $p = .775$, SRMR = 0.02, CFI = 1.00. Results indicate that individuals started training at block 1 with a mean set size of 2.98 ($\mu_i = 2.98$, SE = 0.05, $p < .001$) and significantly increased their performance by 0.49 ($\mu_s = 0.49$, SE = 0.03, $p < .001$), resulting in estimated mean levels of training performance across the five blocks of 2.98 (block 1), 3.47 (block 2), 3.86 (block 3), 4.19 (block 4), and 4.45 (block 5).[3] We found strong evidence for a positive association between the intercept and the slope ($\sigma_{i,s} = 0.03$, SE = 0.01, $p = .004$, $BF_{H1} = 11.98$), suggesting that individuals who showed higher baseline cognitive performance also showed larger training performance gains. Further, there was decisive evidence for individual differences in the variance of baseline cognitive performance ($\sigma^2_i = 0.15$, SE = 0.03, $p < .001$, $BF_{H1} > 100$) and change therein ($\sigma^2_s = 0.03$, SE = 0.01, $p < .001$, $BF_{H1} > 100$).

In the Young-Binding sample, the non-linear baseline LGC model's fit was acceptable, $\chi^2(7) = 23.22$, $p = .002$, SRMR = 0.04, CFI = 0.97. The Young-Binding sample started training at block 1 with a mean set size of 3.46 ($\mu_i = 3.46$, SE = 0.05, $p < .001$) and significantly increased their performance by 0.69 ($\mu_s = 0.69$, SE = 0.04, $p < .001$), resulting in estimated mean levels of training performance across the five blocks of 3.46 (block 1), 4.15 (block 2), 4.62 (block 3), 4.94 (block 4), and 5.19 (block 5). Again, we found decisive evidence for a positive association between the intercept and the slope ($\sigma_{i,s} = 0.05$, SE = 0.01, $p < .001$, $BF_{H1} > 100$), suggesting that

[3] Estimated means are determined by the factor mean of the intercept $\mu_i$ and pattern coefficients $\lambda$ and were computed by the formula: estimated mean = $\mu_i + \lambda * \mu_s$ (see Kline, 2016 for details).

individuals who showed higher baseline cognitive performance also showed larger training performance gains. Further, we found decisive evidence for individual differences in the variance of baseline cognitive performance ($\sigma^2_i$ = 0.12, SE = 0.03, $p < .001$, $BF_{H1} > 100$) and change therein ($\sigma^2_s$ = 0.05, SE = 0.01, $p < .001$, $BF_{H1} > 100$).

Finally, the non-linear baseline LGC model fit the data from the Old-Mixed sample well, $\chi^2$ (7) = 11.83, p = .106, SRMR = 0.05, CFI = 0.99, and showed that older adults started training at block 1 with a mean set size of 3.08 ($\mu_i$ = 3.08, SE = 0.05, $p < .001$) and significantly increased their performance by 0.40 ($\mu_s$ = 0.40, SE = 0.03, $p < .001$), resulting in estimated mean levels of training performance across the five blocks of 3.08 (block 1), 3.48 (block 2), 3.84 (block 3), 4.13 (block 4), and 4.38 (block 5). We found ambiguous evidence for the absence of an association between the intercept and the slope ($\sigma_{i,s}$ = 0.02, SE = 0.01, p = .056, $BF_{H0}$ = 1.39), but again we found decisive evidence for individual differences in the variance of baseline cognitive performance ($\sigma^2_i$ = 0.17, SE = 0.03, $p < .001$, $BF_{H1} > 100$) and change therein ($\sigma^2_s$ = 0.02, SE = 0.00, $p < .001$, $BF_{H1} > 100$).

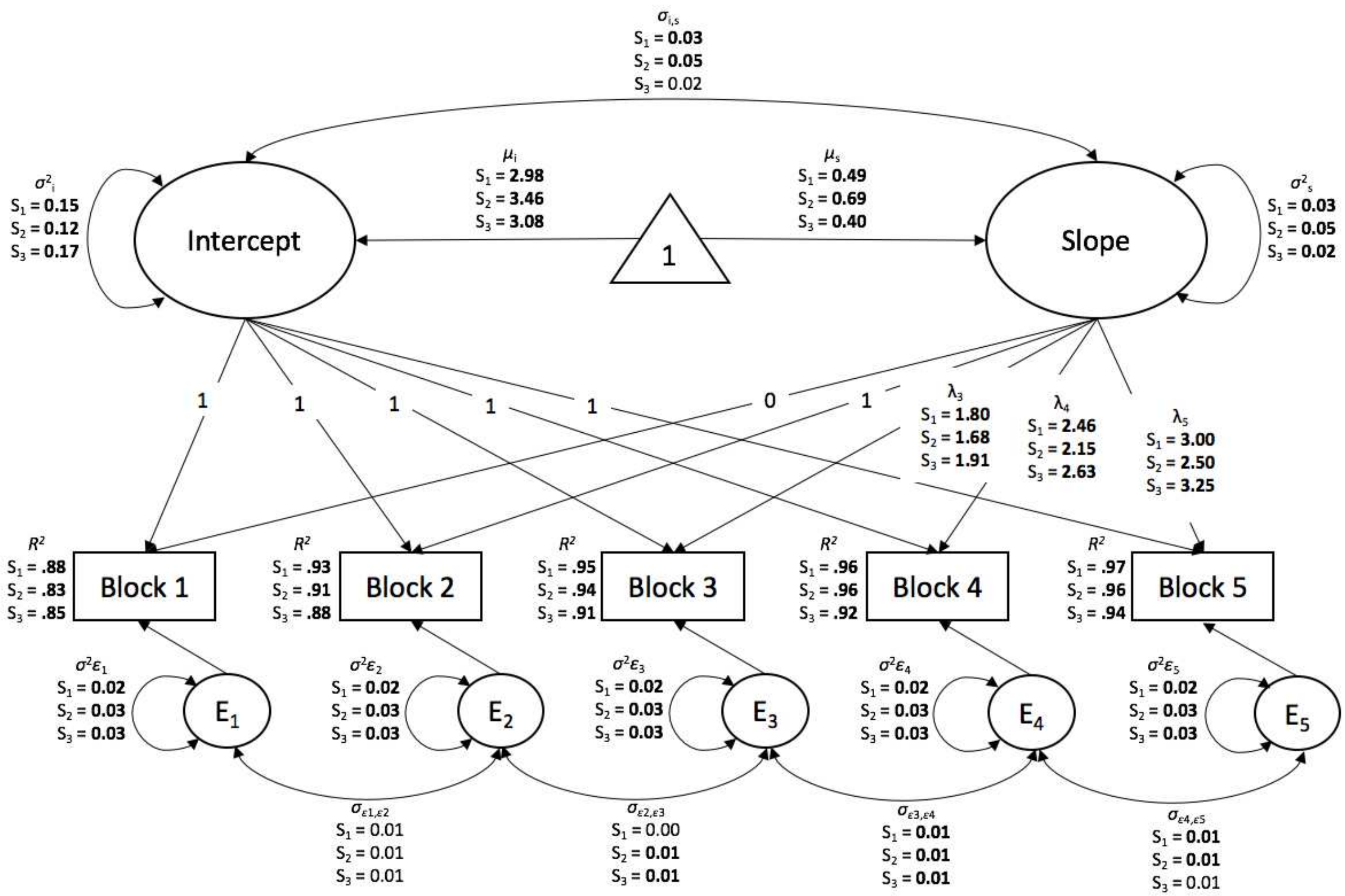



Figure 1. Baseline non-linear latent growth curve model of change in training performance. Bold numbers indicate significance ($p < .05$). Unstandardized estimates are presented for the Young-Updating sample ($S_1$), the Young-Binding sample ($S_2$), and the Old-Mixed sample ($S_3$). Squares represent observed variables (training blocks 1-5),

circles represent latent factors, and the triangle is modeled to represent the means of the latent factors ($\mu_{\text{i}}$ = mean of the intercept, $\mu_{\text{s}}$ = mean of the slope). $\sigma^2_{\text{i}}$ = variance of the intercept; $\sigma^2_{\text{s}}$ = variance of the slope; $\sigma_{\text{i,s}}$ = covariance of intercept and slope; $\lambda_{3\text{-}5}$ = pattern coefficients; $\text{E}_{1\text{-}5}$ = error terms; $\sigma^2\varepsilon_{1\text{-}5}$ = error variances; $\sigma_{\varepsilon,\varepsilon}$ = error covariances.

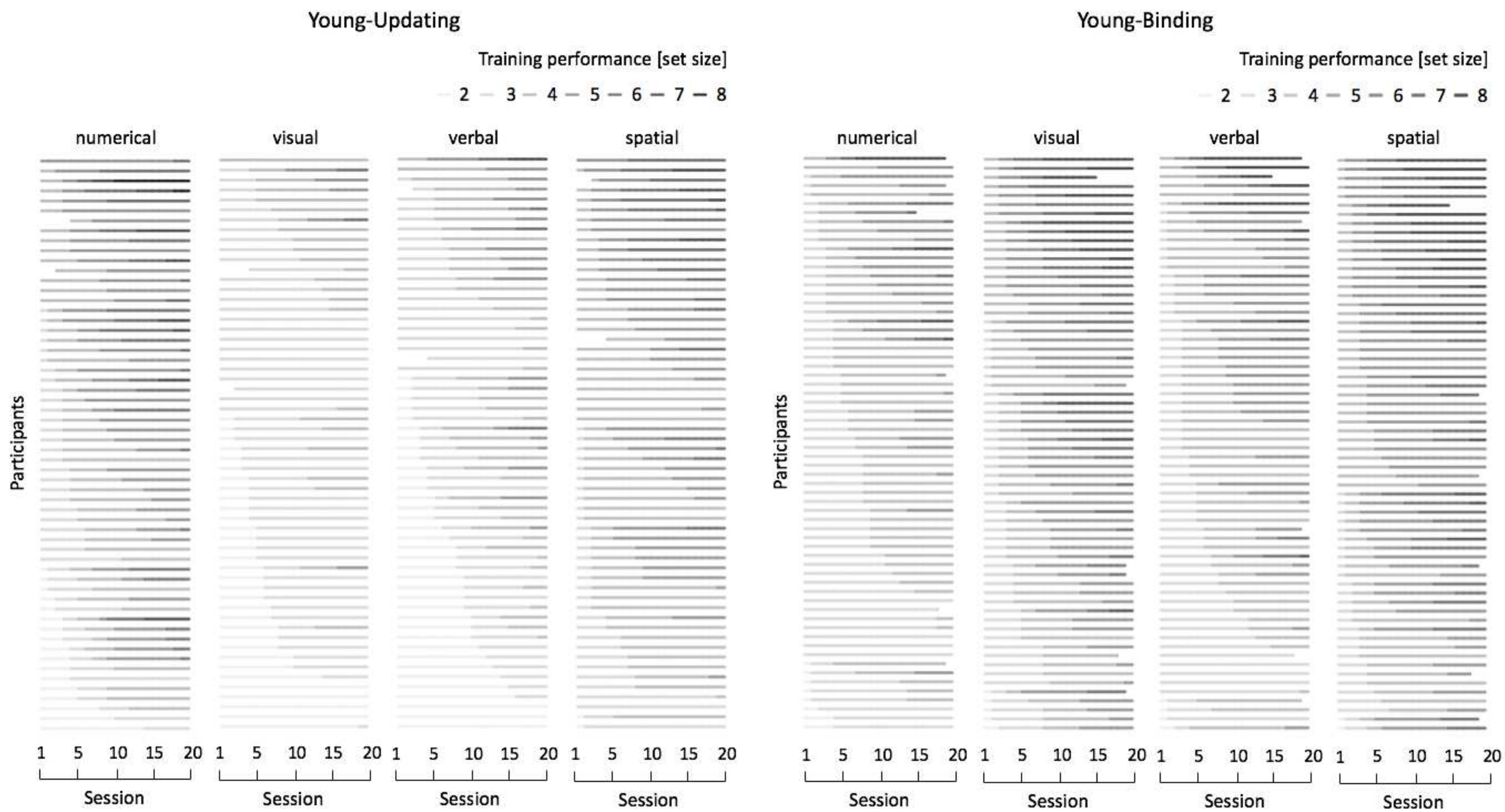



Figure 2. Growth curve plot of task-specific training performance for the Young-Updating and Young-Binding samples. Each line represents an individual, ordered vertically separately for each task using the sorter function implemented in the “longCatEDA” package (Tueller et al., 2016). Shades of grey represent set size achieved at the end of each training session. Thus, lines are darker with increasing training performance and task difficulty.

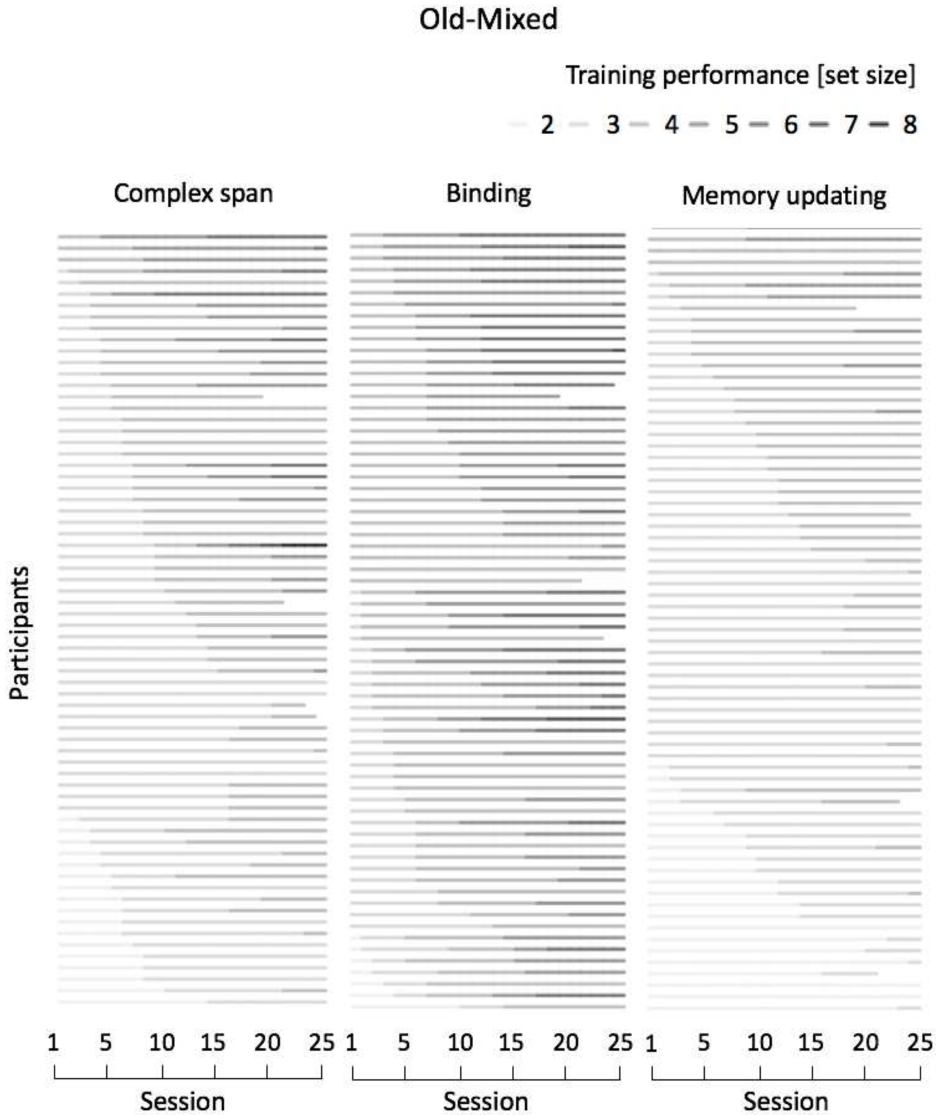
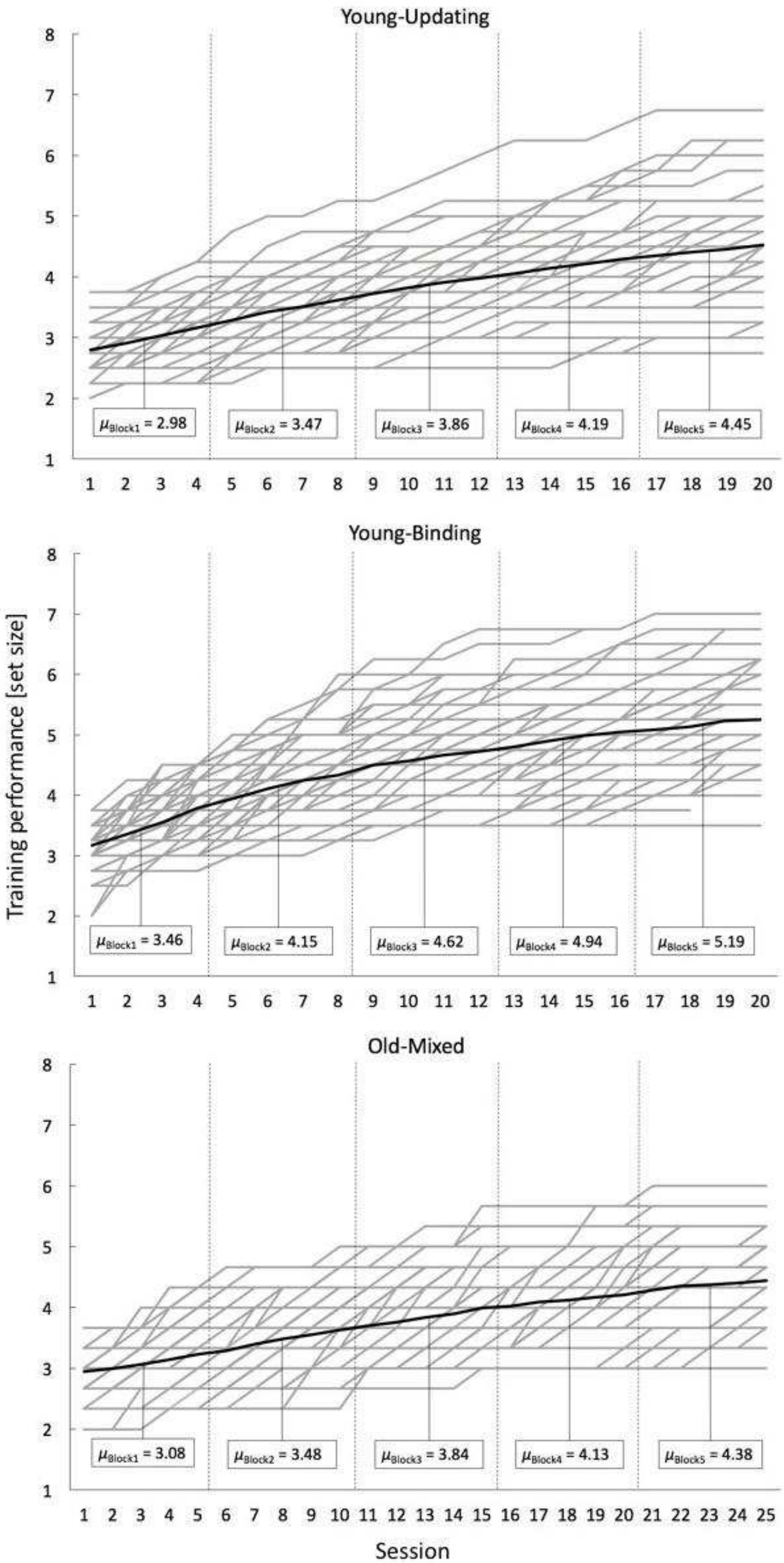



Figure 3. Growth curve plot of task-specific training performance for the Old-Mixed Sample. Each line represents an individual, ordered vertically separately for each task using the sorter function implemented in the “longCatEDA” package (Tueller et al., 2016). Shades of grey represent set size achieved at the end of each training session. Thus, lines are darker with increasing training performance and task difficulty.



Young-Updating
$\mu_{Block1}$ = 2.98
$\mu_{Block2}$ = 3.47
$\mu_{Block3}$ = 3.86
$\mu_{Block4}$ = 4.19
$\mu_{Block5}$ = 4.45
Young-Binding
$\mu_{Block1}$ = 3.46
$\mu_{Block2}$ = 4.15
$\mu_{Block3}$ = 4.62
$\mu_{Block4}$ = 4.94
$\mu_{Block5}$ = 5.19
Old-Mixed
$\mu_{Block1}$ = 3.08
$\mu_{Block2}$ = 3.48
$\mu_{Block3}$ = 3.84
$\mu_{Block4}$ = 4.13
$\mu_{Block5}$ = 4.38
Training performance [set size]
Session

Figure 4. Training performance averaged across training tasks for each individual (grey) and on the group level (black). Estimated means are presented for each training block.

**Association of Individual Differences with Change in Training Performance and Baseline Cognitive Performance**

Descriptive statistics for the individual differences variables are presented in Table 5. To predict training trajectories, we included all variables measuring the same aspect of individual differences simultaneously in the baseline model. Note that although results will be reported separately for the slope and the intercept, the individual differences variables were regressed on both latent factors concurrently.

Table 5

Descriptive Statistics for Individual Differences Variables

| | Sample | | |
|---|---|---|---|
| Individual differences | Young-Updating | Young-Binding | Old-Mixed |
| Demographics | | | |
| Age | 22.57 (2.99) | 24.77 (4.03) | 70.40 (3.72) |
| Gender (f/m) | 39/19 | 45/19 | 30/38 |
| Real-world cognition | | | |
| Education | 5 (0.00) | 5 (0.00) | 5 (1.48) |
| CFQ | - | - | 1.20 (0.42) |
| EPT | - | - | 25.54 (3.05) |
| Motivation | -0.08 (0.95) | 0.09 (0.79) | 5.15 (0.60) |
| Cognition-related beliefs | | | |
| Grit | 2.76 (0.60) | 2.74 (0.61) | 3.74 (0.52) |
| TIS | 4.47 (0.89) | 4.31 (1.01) | 3.98 (1.06) |

| | | | |
|---|---|---|---|
| GSE | 2.98 (0.37) | 3.00 (0.35) | 3.06 (0.37) |
| EXSE | 65.66 (18.22) | 62.84 (17.38) | - |
| NFC | 5.07 (0.69) | 5.03 (0.68) | 5.24 (0.84) |
| Personality | | | |
| Neuroticism | 1.70 (0.63) | 1.60 (0.65) | 1.13 (0.53) |
| Agreeableness | 2.73 (0.60) | 2.81 (0.42) | 2.82 (0.34) |
| Extraversion | 2.40 (0.65) | 2.39 (0.61) | 2.39 (0.50) |
| Openness | 2.73 (0.57) | 2.77 (0.54) | 2.73 (0.43) |
| Conscientiousness | 2.71 (0.58) | 2.75 (0.53) | 2.90 (0.51) |
| Leisure activities | | | |
| Crafts | - | - | 2.31 (1.17) |
| Developmental activities | - | - | 2.41 (0.46) |
| Experiential activities | - | - | 3.40 (0.68) |
| Game playing | - | - | 2.56 (0.89) |
| Physical activities | - | - | 3.13 (0.90) |
| Religious activities | - | - | 2.43 (1.45) |
| Activities with close social partner | - | - | 3.15 (0.55) |
| Group centered public activities | - | - | 1.77 (0.55) |
| Technology use | - | - | 3.14 (0.79) |
| TV watching | - | - | 3.62 (0.90) |
| Travel | - | - | 2.53 (0.57) |
| Training / Computer | | | |
| Computer literacy | - | - | 5.04 (1.52) |
| Training experience (y/n) | - | - | 23/45 |

Note. Values are means and standard deviations in parentheses (median and median absolute deviation in parentheses for education). CFQ = Cognitive Failure Questionnaire; EPT = Everyday Problems Test; TIS = Theories of Intelligence; GSE = General Self-Efficacy scale; EXSE = Self-Efficacy to Regulate Exercise scale; NFC = Need for Cognition.

**Individual differences predicting change in training performance.** Overall, we found only limited evidence for individual differences predicting change in training performance, with most estimates supporting the null hypothesis (see Table 6). There was only one exception. In the Old-Mixed sample, we found substantial evidence for a negative association of growth mindset with change in training performance ($b = -0.37$, $p = .005$, $BF_{H1} = 3.26$), however indicating that individuals who believed more strongly that intelligence is malleable showed less increase in training performance.

For most other individual differences, including demographic variables, real-world cognition, motivation, personality, leisure activities, and computer literacy and training experience, we found evidence against an association with change in training performance, with at least substantial evidence in favor for the null hypothesis ($BF_{H0} \geq 3$).

**Individual differences predicting baseline cognitive performance.** We found some evidence for individual differences predicting baseline cognitive performance, with all evidence, however, being observed in the older adults only (see Table 7).

We found decisive evidence for an association of gender with baseline cognitive performance ($b = 0.45$, $p < .001$, $BF_{H1} > 100$), indicating that male individuals started training at a higher level of performance. Further, there was substantial evidence that age was negatively associated with baseline cognitive performance ($b = -0.32$, $p = .002$, $BF_{H1} = 5.69$), indicating that within the older age group, younger individuals showed higher baseline cognitive performance. Regarding real-world cognition, we found strong evidence for a positive association of EPT performance with baseline cognitive performance ($b = 0.39$, $p < .001$, $BF_{H1} = 18.34$), indicating that individuals who performed better in the EPT also showed higher baseline cognitive performance. In addition, we found substantial evidence for a positive association of grit with baseline cognitive performance ($b = 0.37$, $p = .002$, $BF_{H1} = 6.54$), indicating that grittier individuals showed higher baseline cognitive performance. Regarding personality, we found very strong evidence for a negative association of extraversion with baseline cognitive performance ($b = -0.44$, $p < .001$, $BF_{H1} = 43.40$), indicating that individuals scoring high on extraversion showed lower baseline cognitive performance. Finally, we found substantial evidence for a negative association of religious activities with baseline cognitive performance ($b = -0.34$, $p = .003$, $BF_{H1} = 5.01$), indicating that individuals with high levels of religious activities (e.g., frequent church attendance) started training at a lower level of performance. For most other individual differences, however, we found evidence against an association with baseline cognitive performance, with at least substantial evidence in favor for the null hypothesis ($BF_{H0} \geq 3$).

Table 6

Associations of Individual Differences with Change in Training Performance

| | Young-Updating | | | | Young-Binding | | | | Old-Mixed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Individual differences | b | p | $BF_{H1}$ | $BF_{H0}$ | b | p | $BF_{H1}$ | $BF_{H0}$ | b | p | $BF_{H1}$ | $BF_{H0}$ |
| Demographic variables | | | | | | | | | | | | |
| Age | -0.30 | .014 | 1.61 | 0.62 | -0.26 | .046 | 0.74 | 1.35 | 0.12 | .396 | 0.17 | **5.80** |
| Gender | 0.15 | .244 | 0.25 | **3.98** | 0.27 | .035 | 0.88 | 1.14 | 0.01 | .937 | 0.12 | **8.22** |
| Real-world cognition | | | | | | | | | | | | |
| Education | - | - | - | - | - | - | - | - | 0.31 | .021 | 1.24 | 0.81 |
| CFQ | - | - | - | - | - | - | - | - | 0.07 | .600 | 0.14 | **7.19** |
| EPT | - | - | - | - | - | - | - | - | 0.09 | .511 | 0.15 | **6.66** |
| Motivation | 0.08 | .563 | 0.15 | **6.46** | 0.24 | .058 | 0.63 | 1.59 | -0.13 | .366 | 0.18 | **5.54** |
| Cognition-related beliefs | | | | | | | | | | | | |
| Grit | 0.19 | .138 | 0.37 | 2.71 | 0.11 | .439 | 0.17 | **5.97** | -0.02 | .864 | 0.12 | **8.13** |
| TIS | -0.29 | .028 | 1.06 | 0.95 | -0.16 | .250 | 0.24 | **4.23** | -0.37 | .005 | **3.26** | 0.31 |
| GSE | -0.12 | .467 | 0.17 | **5.87** | -0.20 | .121 | 0.38 | 2.60 | -0.07 | .673 | 0.13 | **7.55** |
| EXSE | -0.11 | .424 | 0.18 | **5.57** | 0.24 | .070 | 0.56 | 1.79 | - | - | - | - |
| NFC | 0.07 | .698 | 0.14 | **7.07** | 0.09 | .562 | 0.15 | **6.77** | 0.05 | .767 | 0.13 | **7.89** |
| Personality | | | | | | | | | | | | |
| Neuroticism | 0.01 | .961 | 0.13 | **7.61** | 0.00 | .978 | 0.12 | **8.00** | -0.13 | .412 | 0.17 | **5.93** |
| Agreeableness | -0.09 | .532 | 0.16 | **6.28** | 0.05 | .683 | 0.14 | **7.37** | 0.12 | .441 | 0.16 | **6.15** |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Extraversion | -0.20 | .196 | 0.29 | **3.44** | -0.29 | .037 | 0.85 | 1.18 | 0.08 | .614 | 0.14 | **7.27** |
| Openness | -0.05 | .688 | 0.14 | **7.03** | 0.04 | .784 | 0.13 | **7.71** | -0.32 | .018 | 1.34 | 0.75 |
| Conscientiousness | -0.27 | .038 | 0.88 | 1.14 | -0.08 | .562 | 0.15 | **6.77** | -0.29 | .055 | 0.65 | 1.54 |
| Leisure activities | | | | | | | | | | | | |
| Crafts | - | - | - | - | - | - | - | - | -0.07 | .637 | 0.14 | **7.38** |
| Developmental activities | - | - | - | - | - | - | - | - | 0.16 | .337 | 0.19 | **5.27** |
| Experiential activities | - | - | - | - | - | - | - | - | -0.09 | .652 | 0.13 | **7.46** |
| Game playing | - | - | - | - | - | - | - | - | 0.05 | .696 | 0.13 | **7.64** |
| Physical activities | - | - | - | - | - | - | - | - | -0.06 | .646 | 0.13 | **7.42** |
| Religious activities | - | - | - | - | - | - | - | - | -0.05 | .703 | 0.13 | **7.67** |
| Activities with social partner | - | - | - | - | - | - | - | - | 0.00 | .992 | 0.12 | **8.24** |
| Public activities | - | - | - | - | - | - | - | - | 0.14 | .380 | 0.18 | **5.66** |
| Technology use | - | - | - | - | - | - | - | - | -0.19 | .193 | 0.27 | **3.68** |
| TV watching | - | - | - | - | - | - | - | - | -0.13 | .352 | 0.19 | **5.40** |
| Travel | - | - | - | - | - | - | - | - | -0.34 | .011 | 1.84 | 0.54 |
| Computer/Training | | | | | | | | | | | | |
| Computer literacy | - | - | - | - | - | - | - | - | -0.28 | .039 | 0.80 | 1.25 |
| Training experience | - | - | - | - | - | - | - | - | 0.05 | .702 | 0.13 | **7.66** |

Note. Bold values represent Bayes factors ≥ 3 indicating substantial evidence for the respective hypothesis. b = standardized estimates; BF = Bayes factor; $H_1$ = alternative hypothesis; $H_0$ = null hypothesis; CFQ = Cognitive Failure Questionnaire; EPT = Everyday Problems Test;

TIS = Theories of Intelligence; GSE = General Self-Efficacy scale; EXSE = Self-Efficacy to Regulate Exercise scale; NFC = Need for Cognition.

Table 7

Associations of Individual Differences with the Baseline Cognitive Performance

| | Young-Updating | | | | Young-Binding | | | | Old-Mixed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Individual differences | b | p | $BF_{H1}$ | $BF_{H0}$ | b | p | $BF_{H1}$ | $BF_{H0}$ | b | p | $BF_{H1}$ | $BF_{H0}$ |
| Demographic variables | | | | | | | | | | | | |
| Age | -0.13 | .336 | 0.21 | **4.86** | -0.27 | .039 | 0.82 | 1.22 | -0.32 | .002 | **5.69** | 0.18 |
| Gender | 0.03 | .815 | 0.13 | **7.41** | 0.17 | .225 | 0.25 | **3.96** | 0.45 | <.001 | **> 100** | 0.01 |
| Real-world cognition | | | | | | | | | | | | |
| Education | - | - | - | - | - | - | - | - | 0.25 | .030 | 1.00 | 1.00 |
| CFQ | - | - | - | - | - | - | - | - | -0.09 | .429 | 0.17 | **6.06** |
| EPT | - | - | - | - | - | - | - | - | 0.39 | <.001 | **18.34** | 0.05 |
| Motivation | 0.18 | .179 | 0.31 | **3.25** | 0.20 | .127 | 0.37 | 2.71 | -0.13 | .325 | 0.19 | **5.15** |
| Cognition-related beliefs | | | | | | | | | | | | |
| Grit | 0.03 | .791 | 0.14 | **7.35** | 0.20 | .129 | 0.37 | 2.71 | 0.37 | .002 | **6.54** | 0.15 |
| TIS | -0.34 | .007 | 2.72 | 0.37 | 0.16 | .263 | 0.23 | **4.37** | -0.06 | .635 | 0.14 | **7.37** |
| GSE | 0.00 | .997 | 0.13 | **7.61** | -0.09 | .498 | 0.16 | **6.38** | -0.29 | .033 | 0.95 | 1.06 |
| EXSE | 0.14 | .288 | 0.23 | **4.41** | 0.23 | .080 | 0.51 | 1.97 | - | - | - | - |
| NFC | 0.23 | .149 | 0.35 | 2.86 | 0.15 | .310 | 0.21 | **4.86** | 0.12 | .420 | 0.17 | **5.99** |
| Personality | | | | | | | | | | | | |
| Neuroticism | -0.03 | .823 | 0.13 | **7.43** | 0.06 | .657 | 0.14 | **7.26** | -0.28 | .021 | 1.31 | 0.76 |
| Agreeableness | -0.16 | .274 | 0.23 | **4.28** | 0.23 | .072 | 0.55 | 1.83 | 0.20 | .090 | 0.47 | 2.15 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Extraversion | 0.11 | .504 | 0.16 | **6.11** | -0.18 | .213 | 0.26 | **3.81** | -0.44 | <.001 | **43.40** | 0.02 |
| Openness | -0.02 | .868 | 0.13 | **7.51** | 0.15 | .247 | 0.24 | **4.21** | -0.04 | .722 | 0.13 | **7.74** |
| Conscientiousness | -0.15 | .292 | 0.22 | **4.46** | -0.03 | .833 | 0.13 | **7.83** | 0.32 | .007 | 2.94 | 0.34 |
| Leisure Activities | | | | | | | | | | | | |
| Crafts | - | - | - | - | - | - | - | - | 0.25 | .046 | 0.75 | 1.33 |
| Developmental activities | - | - | - | - | - | - | - | - | 0.24 | .085 | 0.49 | 2.05 |
| Experiential activities | - | - | - | - | - | - | - | - | -0.31 | .061 | 0.62 | 1.62 |
| Game playing | - | - | - | - | - | - | - | - | 0.08 | .514 | 0.15 | **6.68** |
| Physical activities | - | - | - | - | - | - | - | - | -0.03 | .838 | 0.12 | **8.07** |
| Religious activities | - | - | - | - | - | - | - | - | -0.34 | .003 | **5.01** | 0.20 |
| Activities with social partner | - | - | - | - | - | - | - | - | -0.09 | .490 | 0.15 | **6.51** |
| Public activities | - | - | - | - | - | - | - | - | 0.21 | .134 | 0.35 | 2.83 |
| Technology use | - | - | - | - | - | - | - | - | 0.08 | .563 | 0.14 | **6.98** |
| TV watching | - | - | - | - | - | - | - | - | 0.07 | .572 | 0.14 | **7.03** |
| Travel | - | - | - | - | - | - | - | - | 0.03 | .838 | 0.12 | **8.07** |
| Computer/Training | | | | | | | | | | | | |
| Computer literacy | - | - | - | - | - | - | - | - | 0.20 | .114 | 0.39 | 2.57 |
| Training experience | - | - | - | - | - | - | - | - | 0.17 | .173 | 0.29 | **3.41** |

Note. Bold values represent Bayes factors ≥ 3 indicating substantial to decisive evidence for the respective hypothesis. b = standardized estimates; BF = Bayes factor; $H_1$ = alternative hypothesis; $H_0$ = null hypothesis; CFQ = Cognitive Failure Questionnaire; EPT = Everyday

Problems Test; TIS = Theories of Intelligence; GSE = General Self-Efficacy scale; EXSE = Self-Efficacy to Regulate Exercise scale; NFC = Need for Cognition.

**Additional Analyses of the First Training Block**

A limitation of our modeling approach is that the intercept represents the mean performance across the first block (i.e., the average set size of the first 4 or 5 training sessions, depending on the sample). Thus, this analysis does not allow to directly predict change in training performance during this first training block in the context of overall change in training performance. Therefore, to investigate how individual differences are associated with baseline cognitive performance at the first training session and change in training performance across the first training block, we additionally ran the same models for the first training block only, with the first training session as the intercept and change modeled across the first four to five training sessions, depending on the sample. Detailed results of these analyses are reported in the supplemental material (see Tables S3 to S6, Figure S1).

Overall, although the BFs were somewhat lower in these additional analyses (possibly due to the increased noise in the non-averaged data), the pattern of results was largely similar to the findings of our primary analyses, with a few exceptions. Whereas a model assuming a non-linear change in training performance still fitted the data of the Old-Mixed sample best, nested model comparisons indicated the best fit for a model assuming a linear change in both younger samples (see Table S3 in the supplemental material). Hence, younger, but not older adults showed steeper performance increases during the first few sessions than across all sessions. As for the primary analyses, evidence for the variance of baseline cognitive performance and change in cognitive performance was decisive for all samples (see Table S4 in the supplemental material). However, different to the primary analyses, we found substantial evidence for the absence of an association between the intercept and slope in both younger samples. The evidence for this association was again ambiguous for the older adults (see Table S4 in the supplemental material).

Similar to the primary analyses, most predictors were also unrelated to change in training performance over the first few training sessions (see Table S5 in the supplemental material). In addition to the now strong evidence for a negative association with growth mindset (b = -0.44, p = .001, $BF_{H1}$ = 10.37), we found substantial evidence for a negative association with age (b = -0.36, p = .004, $BF_{H1}$ = 3.38), indicating that, within the older sample, younger individuals changed more during the first training block. Taken together with the finding that the slope followed a linear function in the younger samples, but a non-linear function in the older sample, this suggests that age differences play a bigger role at the beginning of training than at later stages.

Results were also largely similar for the predictors of baseline cognitive performance at the first session, with a few exceptions (see Table S6 in the supplemental material). First, in the Old-Mixed sample, there was substantial evidence for a negative association of general self-efficacy with performance in the first session (b = -0.39, p = .001, $BF_{H1}$ = 7.03). Second, in the Young-Updating sample, we found substantial evidence for a negative association of a growth mindset (b = -0.38, p = .002, $BF_{H1}$ = 5.35). Third, the associations of the intercept with age and religious activities were no longer substantial when analysing only the first session.

## Discussion

The objectives of the present work were threefold. First, we estimated individual training trajectories. Second, we related baseline cognitive performance (i.e., the intercept) to change in training performance across the training phase (i.e., the slope). Third, we examined the extent to which individual differences were predictive of change in training performance. We modeled LGCs for three WM training interventions in younger and older adults that comprised a broad set of potential individual differences variables previously discussed in the literature, including demographic variables, motivation, cognition-related beliefs, and personality traits. Using BFs enabled us to evaluate the strength of evidence for the presence as well as the absence of a possible association between individual differences in the above variables and change in training performance.

Performance improved non-linearly across the training phase in all three samples. In line with the magnification account, this change in training performance was positively associated with baseline cognitive performance, indicating that individuals who started off on higher performance levels also improved more throughout the training phase. However, whereas evidence for the presence of this relationship was strong to decisive in the two younger samples, we found ambiguous evidence for the absence of it in the older sample. Finally, although baseline cognitive performance was predicted by individual differences in some variables (i.e., demographics, real-world cognition, cognition-related beliefs, personality, and leisure activities), only 1 out of 29 variables predicted change in training performance, and did so only inconsistently across samples. More specifically, we found that, in the older sample, growth mindset was negatively associated with change in training performance. Taken together, our findings suggest that changes observed during training are best predicted by baseline cognitive performance, with individual differences in demographic variables, real-world cognition, motivation, cognition-related beliefs, personality traits, leisure activities, and computer and training experience playing a negligible role only.

**Magnification of Training Performance**

In all three samples, individuals substantially increased their performance across the training phase, with a steeper increase at the beginning of the training phase leveling off toward the end of the training phase. Large training effects are an established finding in the literature across various training regimes in both younger (e.g., Brehmer et al., 2012; Jaeggi et al., 2008; Sprenger et al., 2013; von Bastian & Oberauer, 2013) and older adults (e.g., von Bastian et al., 2013; Zimmermann, von Bastian, Röcke, Martin, & Eschen, 2016; see Karbach & Verhaeghen, 2014 for a meta-analysis) indicating that improvements in complex cognitive tasks are not limited to younger adults, but extend into old age.

The positive association between baseline cognitive performance and change in training performance is in line with studies reporting that general WM performance strongly predicts cognitive learning in associative and category-learning tasks (e.g., Lewandowsky, 2011; Tamez, Myerson, & Hale, 2012) and previous literature on age-related and ability-related magnification effects in the context of cognitive training (e.g., Bürki et al., 2014; Schmiedek et al., 2010). Magnification effects are more typically observed in the context of strategy-based training than process-based training (e.g., Karbach & Verhaeghen, 2014), possibly indicating that the training intervention in this study facilitated strategy acquisition (for a more detailed discussion, see De Simoni & von Bastian, 2017; Guye & von Bastian, 2017). It has been argued that individuals with higher levels of cognitive performance at baseline have more cognitive capacity available to acquire and perform strategies to enhance cognitive efficiency during training (Lövdén et al., 2012). However, the positive association between baseline cognitive performance and change in cognitive performance was less pronounced in the older sample, providing ambiguous evidence for the absence of this association in the older adults. One possible explanation for this finding is that, although often proclaimed otherwise, older adults in our sample differed somewhat less than younger adults in their training slope ($\sigma^2_s$ = 0.02 compared to $\sigma^2_s$ = 0.05 in the Young-Binding and $\sigma^2_s$ = 0.03 in the Young-Updating samples). Hence, it is possible that power was simply too low to detect the positive relationship, as indicated by the ambiguous BF. Furthermore, future studies are needed to directly compare the association of baseline cognitive ability with change in cognitive performance in younger and older adults in order to draw conclusions regarding age-related differences in magnification effects.

**Limited Evidence for Individual Differences Predicting Change in Training Performance**

Concerning the debate about the effectiveness of cognitive training interventions, an often-voiced explanation for inconsistencies between the studies is the potential role of individual differences on training outcomes (e.g., Shah et al., 2012), with individually-tailored interventions potentially maximizing the effects of cognitive training. We indeed found substantial variance among individuals in change of training performance in all samples that could be potentially predicted by variables that had been discussed in the past (Katz et al., 2016). Therefore, we examined how (1) demographic variables, (2) real-world cognition, (3) motivation, (4) cognition-related beliefs, (5) personality, (6) leisure activities, and (7) computer literacy and training experience predicted variance in the training trajectories. Based on previous literature, we expected a positive association of motivation, growth mindset, and conscientiousness, and a negative association of age with change in training performance. For all the other individual differences, the analyses were exploratory. However, our results did not support our expectations.

First, we found substantial evidence for the absence of an association of age with change in training performance across the entire training intervention at least in the older sample. However, in our additional analyses we found substantial evidence for a positive association of age with change in training performance in the first training block for older adults, indicating that age differences might be relevant during early stages of training, but less so later on. In addition, change in training performance was positively associated with baseline performance, implying that age and initial cognitive performance indeed may need to be conceptually separated when examining magnification and compensation effects (von Bastian & Oberauer, 2014).

Second, we found evidence for the absence of an association of change in training performance with previously proposed personality traits such as neuroticism and conscientiousness. Hence, although neuroticism has been reported to be associated with mean training performance and transfer effects (e.g., Studer-Luethi et al., 2012; 2016), it may only play a neligible role in predicting change in training performance. This is in line with previous findings showing no significant association of neuroticism with training gains (Studer-Luethi et al., 2012; 2016).

Third, we found evidence for the absence of an association of training-related motivation with change in training performance. Although previous literature has shown within-person associations between daily motivation and daily cognitive performance during a training intervention (Brose et al., 2012), we did not observe such a relationship on the between-person level, suggesting that motivation might be more strongly linked to daily fluctuations in cognitive performance than to overall training trajectories.

Fourth and contrary to our expectations, we found evidence for a negative association of growth mindset with change in training performance in the older sample. Similarly, Thompson and colleagues (2013) reported a marginally significant negative association of growth mindset with improvements in a trained WM task in younger adults. We can only speculate about what causes this rather counterintuitive finding, but one possible explanation could be that individuals with high levels of growth mindset are so heavily focused on changing their cognitive performance that they pay too much attention to their cognitive performance, drawing away resources that would be necessary to perform the training tasks efficiently (see also Studer-Luethi et al., 2012).

**Limitations**

Despite several strengths of the present study, there are some limitations. First, our analyses do not allow for a direct comparison between the three samples. Although they were all undergoing highly similar training regimes, there were slight differences between the interventions regarding the exact tasks being used in the different age-groups (single vs. mixed-paradigm training), and the features of the training interventions (e.g., frequency of the training sessions, monetary reward). Thus, in order to directly compare the presence or absence of the individual differences in younger and older adults, future studies should pursue an age-comparative approach.

Second, the averaging across several training tasks and training sessions to improve the robustness of our performance indicators, was, unavoidably, accompanied some shortcomings. First, averaging across multiple sessions and tasks comes with a loss of more fine-grained information regarding the performance in the single tasks and sessions. Second, it prevented us from predicting early performance changes in context of overall change in training performance (i.e., the first 4 or 5 sessions, but see supplemental material). Using the average across the first few sessions as a measure of baseline cognitive performance comes, however, also with the advantage to reduce noise from two sources of unwanted variance, that is (1) from training-specific adjustment processes at the beginning of the training (i.e., getting used to the computer, understanding the nature of the training tasks), and (2) from substantial day-to-day variability in cognitive performance (Schmiedek, Lövdén, & Lindenberger, 2013).

Finally, although our group sizes were considerably larger than the median group size in the cognitive training literature (n = 22; Lampit et al., 2014), they are still fairly small when using SEM and relying on traditional NHST. In the presence of small sample sizes, p-values can vary greatly, known as “the dance of the p-values” (Bogg & Lasecki, 2015; Cumming, 2011; Halsey, Curran-Everett, Vowler, & Drummond, 2015; von Bastian et al., 2017). To overcome this limitation, we additionally evaluated the evidence for and against the existence of links between the individual differences variables and change in training performance using BFs, as they vary less when power is low (Dienes, 2014). The size of the BFs indicate that our sample sizes were sufficient to provide conclusive evidence for the absence of the majority of investigated associations.

## Conclusion

To the best of our knowledge, our study was the first to comprehensively investigate a broad range of individual differences in cognitive lab and real-world performance, demographics, motivation, cognition-related beliefs, personality traits, leisure activities, as well as computer literacy and training experience, which had previously been discussed to potentially predict change in training performance, in different study populations (i.e., younger and older adults). However, although we found some of the proposed variables predicted baseline cognitive performance, change in training performance was predicted primarily by baseline cognitive performance in the younger adults, suggesting that individuals scoring higher in the beginning of training also showed more pronounced improvements across the training phase.

References

Ackerman, P. L., & Lohman, D. F. (2006). Individual Differences in Cognitive Functions. In P. A. Alexander & P. H. Winne (Eds.), Handbook of Educational Psychology (2nd ed., pp. 139–161). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., … Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. Frontiers in Psychology, 5, 398. http://doi.org/10.3389/fpsyg.2014.00398

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. Psychonomic Bulletin & Review, 22(2), 366–377. http://doi.org/10.3758/s13423-014-0699-x

Baker, T. J., & Bichsel, J. (2006). Personality predictors of intelligence: Differences between young and cognitively healthy older adults. Personality and Individual Differences, 41(5), 861–871. http://doi.org/10.1016/j.paid.2006.02.017

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), Self-Efficacy Beliefs of Adolescents (pp. 307–337). Greenwich, CT: Information Age Publishing.

Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., & Becic, E. (2008). Transfer effects in task-set cost and dual-task cost after dual-task training in older and younger adults: further evidence for cognitive plasticity in attentional control in late adulthood. Experimental Aging Research, 34(3), 188–219. http://doi.org/10.1080/03610730802070068

Bogg, T., & Lasecki, L. (2015). Reliable gains? Evidence for substantially underpowered designs in studies of working memory training transfer to fluid intelligence. Frontiers in

Psychology, 5. http://doi.org/10.3389/fpsyg.2014.01589

Borella, E., Carretti, B., Cantarella, A., Riboldi, F., Zavagnin, M., & De Beni, R. (2014). Benefits of training visuospatial working memory in young-old and old-old. Developmental Psychology, 50(3), 714–727. http://doi.org/10.1037/a0034293

Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: Evidence of transfer and maintenance effects. Psychology and Aging, 25(4), 767–778. http://doi.org/10.1037/a0020683

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. Frontiers in Human Neuroscience, 6. http://doi.org/10.3389/fnhum.2012.00063

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. British Journal of Clinical Psychology, 21(1), 1–16. http://doi.org/10.1111/j.2044-8260.1982.tb01421.x

Brose, A., Schmiedek, F., Lövdén, M., & Lindenberger, U. (2012). Daily variability in working memory is coupled with negative affect: The role of attention and motivation. Emotion, 12(3), 605–617. http://doi.org/10.1037/a0024436

Buitenweg, J. I. V., Murre, J. M. J., & Ridderinkhof, K. R. (2012). Brain training in progress: a review of trainability in healthy seniors. Frontiers in Human Neuroscience, 6, 183. http://doi.org/10.3389/fnhum.2012.00183

Bürki, C., Ludwig, C., Chicherio, C., & de Ribaupierre, A. (2014). Individual differences in cognitive plasticity: an investigation of training curves in younger and older adults. Psychological Research, 78(6), 821–835. http://doi.org/10.1007/s00426-014-0559-3

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. Journal of Personality and Social

Psychology, 42(1), 116–131. http://doi.org/10.1037/0022-3514.42.1.116

Chamorro-Premuzic, T., Furnham, A., & Ackerman, P. L. (2006). Ability and personality correlates of general knowledge. Personality and Individual Differences, 41(3), 419–429. http://doi.org/10.1016/j.paid.2005.11.036

Clark, C. M., Lawlor-Savage, L., & Goghari, V. M. (2017). Working memory training in healthy young adults : Support for the null from a randomized comparison to active and passive control groups. PLoS ONE, 12(5), 1–25. http://doi.org/10.1371/journal.pone.0177707 May

Costa, P. T., & McCrae, R. R. (1992). Revised NEO personality inventory (NEO PI-R) and NEP five-factor inventory (NEO-FFI): professional manual. Odessa, FL: Psychological Assessment Resources.

Cumming, G. (2011). Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. New York: Routledge.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. Journal of Cognitive Development, 11(2), 121–136. http://doi.org/10.1080/15248371003699969

De Simoni, C., & von Bastian, C. C. (manuscript submitted for publication). Process-specific effects of working memory training.

Deci, E. L., & Ryan, R. M. (2016). Intrinsic Motivation Inventory. Retrieved from http://selfdeterminationtheory.org/intrinsic-motivation-inventory/

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. Frontiers in Psychology, 5. http://doi.org/10.3389/fpsyg.2014.00781

Dougherty, M. R., Hamovitz, T., & Tidwell, J. W. (2016). Reevaluating the effectiveness of n-back training on transfer through the Bayesian lens: Support for the null. Psychonomic

Bulletin & Review, 23(1), 306–316. http://doi.org/10.3758/s13423-015-0865-9

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. Journal of Personality and Social Psychology, 92(6), 1087–1101. http://doi.org/10.1037/0022-3514.92.6.1087

Dweck, C. S. (2000). Self-theories: Their role in motivation, personality, and development. Philadelphia, PA: Psychology Press.

Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., & Strobel, A. (2010). Same or different? Clarifying the relationship of need for cognition to personality and intelligence. Personality and Social Psychology Bulletin, 36(1), 82–96. http://doi.org/10.1177/0146167209351886

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12(3), 189–198. http://doi.org/10.1016/0022-3956(75)90026-6

Graham, E. K., & Lachman, M. E. (2012). Personality stability is associated with better cognitive performance in adulthood: Are the stable more able? Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 67(5), 545–554. http://doi.org/10.1093/geronb/gbr149

Grimm, K. J., Ram, N., & Estabrook, R. (2017). Growth Modeling: Structural Equation and Multilevel Modeling Approaches. New York: Guilford Press.

Guye, S., Röcke, C., Mérillat, S., von Bastian, C. C., & Martin, M. (2016). Adult Lifespan. In T. Strobach & J. Karbach (Eds.), Cognitive Training: An Overview of Features and Applications (pp. 45–55). Berlin: Springer.

Guye, S., & von Bastian, C. C. (in press). Working Memory Training in Older Adults: Evidence for the Absence of Transfer. Psychology and Aging. doi: http://dx.doi.org/10.1037/pag0000206

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. Nature Methods, 12(3), 179–185. http://doi.org/10.1038/nmeth.3288

Hill, B. D., Foster, J. D., Elliott, E, M., Shelton, J. T., McCain, J., & Gouvier, W. D. (2013). Need for cognition is related to higher general intelligence, fluid intelligence, and crystallized intelligence, but not working memory. Journal of Research in Personality, 47(1), 22–25. http://doi.org/10.1016/j.jrp.2012.11.001

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. http://doi.org/10.1080/10705519909540118

Ishihara, S. (1917). Tests for Colour Blindness. Tokyo: Handaya Hongo Harukichi.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. Proceedings of the National Academy of Sciences of the United States of America, 105(19), 6829–6833. http://doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. Memory & Cognition, 42(3), 464–480. http://doi.org/10.3758/s13421-013-0364-z

Jopp, D., & Hertzog, C. (2007). Activities, self-referent memory beliefs, and cognitive performance: evidence for direct and mediated relations. Psychology and Aging, 22(4), 811–825. http://doi.org/10.1037/0882-7974.22.4.811

Jopp, D., & Hertzog, C. (2010). Assessing Adult Leisure Activities: An Extension of a Self-Report Activity Questionnaire. Psychological Assessment, 22(1), 108–120. http://doi.org/10.1037/a0017662

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. Developmental Science, 12(6), 978–990. http://doi.org/10.1111/j.1467-7687.2009.00846.x

Karbach, J., & Verhaeghen, P. (2014). Making Working Memory Work: A Meta-Analysis of Executive-Control and Working Memory Training in Older Adults. Psychological Science, 25(11), 2027–2037. http://doi.org/10.1177/0956797614548725

Katz, B., Jones, M. R., Shah, P., Buschkuehl, M., & Jaeggi, S. M. (2016). Individual Differences and Motivational Effects in Cognitive Training Research. In T. Strobach & J. Karbach (Eds.), Cognitive Training: An Overview of Features and Applications (pp. 157–166). Berlin: Springer.

Kelly, M. E., Loughrey, D., Lawlor, B. A., Robertson, I. H., Walsh, C., & Brennan, S. (2014). The impact of cognitive training and mental stimulation on cognitive and everyday functioning of healthy older adults: A systematic review and meta-analysis. Ageing Research Reviews, 15, 28–43. http://doi.org/10.1016/j.arr.2014.02.004

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. Sociological Methods & Research, 44(3), 486–507. http://doi.org/10.1177/0049124114543236

Kirk, N. W., Fiala, L., Scott-Brown, K. C., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. Journal of Cognitive Psychology, 26(6), 640–648. http://doi.org/10.1080/20445911.2014.929580

Kliegl, R., Smith, J., & Baltes, P. B. (1990). On the locus and process of magnification of age differences during mnemonic training. Developmental Psychology, 26(6), 894–904. http://doi.org/10.1037/0012-1649.26.6.894

Kline, R. B. (2016). Mean Structures and Latent Growth Models. In Principles and Practice of Structural Equation Modeling (4th ed., pp. 369–393). New York, NY: Guilford Press.

Könen, T., & Karbach, J. (2015). The benefits of looking at intraindividual dynamics in cognitive training data. Frontiers in Psychology, 6, 615. http://doi.org/10.3389/fpsyg.2015.00615

Lampit, A., Hallock, H., & Valenzuela, M. (2014). Computerized Cognitive Training in Cognitively Healthy Older Adults: A Systematic Review and Meta-Analysis of Effect Modifiers. PLoS Medicine, 11(11), e1001756. http://doi.org/10.1371/journal.pmed.1001756

Lewandowsky, S. (2011). Working memory capacity and categorization: individual differences and modeling. Journal in Experimental Psychology: Learning, Memory, and Cognition, 37(3), 720–738. http://doi.org/10.1037/a0022639

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. Behavior Research Methods, 42(2), 571–585. http://doi.org/10.3758/BRM.42.2.571

Lövdén, M., Brehmer, Y., Li, S.-C., & Lindenberger, U. (2012). Training-induced compensation versus magnification of individual differences in memory performance. Frontiers in Human Neuroscience, 6, 141. http://doi.org/10.3389/fnhum.2012.00141

Melby-Lervåg, M., & Hulme, C. (2013). Is Working Memory Training Effective? A Meta-Analytic Review. Developmental Psychology, 49(2), 270–291. http://doi.org/10.1037/a0028228

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of "Far Transfer": Evidence from a Meta-Analytic Review. Perspectives on Psychological Science, 11(4), 512–534. http://doi.org/10.3837/tiis.0000.00.000

Minear, M., Brasher, F., Brandt Guerrero, C., Brasher, M., Moore, A., & Sukeena, J. (2016). A simultaneous examination of two forms of working memory training: Evidence for near transfer only. Memory & Cognition, 44(7), 1014–1037. http://doi.org/10.3758/s13421-016-0616-9

Moreau, D., Kirk, I. J., & Waldie, K. E. (2016). Seven pervasive statistical flaws in cognitive training interventions. Frontiers in Human Neuroscience, 10, 3. http://doi.org/10.3389/fnhum.2016.00153

Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: Gauging the generality of cognitive intervention effects. Restorative Neurology and Neuroscience, 27(5), 435–453. http://doi.org/10.3233/RNN-2009-0496

Oberauer, K. (2005). Binding and Inhibition in Working Memory: Individual and Age Differences in Short-Term Recognition. Journal of Experimental Psychology: General, 134(3), 368–387. http://doi.org/10.1037/0096-3445.134.3.368

Paap, K. R., Johnson, H. R., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? Journal of Cognitive Psychology, 26(6), 615–639. http://doi.org/10.1080/20445911.2014.944914

Rabbitt, P., Diggle, P., Holland, F., & McInnes, L. (2004). Practice and drop-out effects during a 17-year longitudinal study of cognitive aging. Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 59(2), 84–97.

http://doi.org/10.1093/geronb/59.2.P84

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.r-project.org

Rheinberg, F., Vollmeyer, R., & Bruns, B. D. (2001). FAM: EIn Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen. Diagnostica, 47(2), 57–66. http://doi.org/10.1026//0012-1924.47.2.57

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2). http://doi.org/10.18637/jss.v048.i02

Schaie, K. W., Willis, S. L., & Caskie, G. I. L. (2004). The Seattle longitudinal study: relationship between personality and cognition. Neuropsychology, Development, and Cognition Section B: Aging, Neuropsychology, and Cognition, 11(2–3), 304–324. http://doi.org/10.1080/13825580490511134

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit Measures. Methods of Psychological Research Online, 8(2), 23–74. http://doi.org/10.1002/0470010940

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred Days of Cognitive Training Enhance Broad Cognitive Abilities in Adulthood: Findings from the COGITO Study. Frontiers in Aging Neuroscience, 2(27), 1–10. http://doi.org/10.3389/fnagi.2010.00027

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2013). Keeping It Steady: Older Adults Perform More Consistently on Cognitive Tasks Than Younger Adults. Psychological Science, 24(9), 1747–1754. http://doi.org/10.1177/0956797613479611

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2014). A task is a task is a task: Putting complex span, n-back, and other working memory indicators in psychometric context.

Frontiers in Psychology, 5. http://doi.org/10.3389/fpsyg.2014.01475

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does Working Memory Training Transfer? A Meta-Analysis Including Training Conditions as Moderators. Educational Psychologist, 50(2), 138–166. http://doi.org/10.1080/00461520.2015.1036274

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in Health Psychology: A User's Po*rtfolio. Causal and Control Beliefs (pp. 35–37). Windsor, UK: NFER-NELSON.

Shah, P., Buschkuehl, M., Jaeggi, S., & Jonides, J. (2012). Cognitive training for ADHD: The importance of individual differences. Journal of Applied Research in Memory and Cognition, 1(3), 204–205. http://doi.org/10.1016/j.jarmac.2012.07.001

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent Evidence and Development of a Shorter Version. Clinical Gerontologist, 5(1–2), 165–173. http://doi.org/10.1300/J018v05n01

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? Psychological Bulletin, 138(4), 628–654. http://doi.org/10.1037/a0027473

Soubelet, A., & Salthouse, T. A. (2011). Personality- cognition relations across adulthood. Developmental Psychology, 47(2), 303–310. http://doi.org/10.1037/a0021816

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. Psychonomic Bulletin & Review. http://doi.org/10.3758/s13423-016-1217-0

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. R., … Dougherty, M. R. (2013). Training working memory: Limits of transfer. Intelligence, 41(5), 638–663. http://doi.org/10.1016/j.intell.2013.07.013

Stern, E. (2017). Individual differences in the learning potential of human beings. Npj Science of Learning, 2(1). http://doi.org/10.1038/s41539-016-0003-0

Studer-Luethi, B., Bauer, C., & Perrig, W. J. (2016). Working memory training in children: Effectiveness depends on temperament. Memory & Cognition, 44(2), 171–186. http://doi.org/10.3758/s13421-015-0548-9

Studer-Luethi, B., Jaeggi, S. M., Buschkuehl, M., & Perrig, W. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. Personality and Individual Differences, 53(1), 44–49. http://doi.org/10.1016/j.paid.2012.02.012

Tamez, E., Myerson, J., & Hale, S. (2012). Contributions of associative learning to age and individual differences in fluid intelligence. Intelligence, 40(5), 518–529. http://doi.org/10.1016/j.intell.2012.04.004

Thompson, T. W., Waskom, M. L., Garel, K.-L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., … Gabrieli, J. D. E. (2013). Failure of Working Memory Training to Enhance Cognition or Intelligence. PLoS ONE, 8(5), e63614. http://doi.org/10.1371/journal.pone.0063614

Titz, C., & Karbach, J. (2014). Working memory and executive functions: effects of training on academic achievement. Psychological Research, 78(6), 852–868. http://doi.org/10.1007/s00426-013-0537-1

Tueller, S. J., Van Dorn, R. A., & Bobashev, G. V. (2016). Visualization of Categorical Longitudinal and Times Series Data. Methods Report RTI Press, 1–22. http://doi.org/10.3768/rtipress.2016.mr.0033.1602

Unsworth, N., Redick, T. S., McMillan, B. D., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2015). Is Playing Video Games Related to Cognitive Abilities? Psychological Science, 26(6), 759–774. http://doi.org/10.1177/0956797615570367

Urbánek, T., & Marček, V. (2015). Investigating the effectiveness of working memory training in the context of Personality Systems Interaction theory. Psychological Research, 80(5), 877–888. http://doi.org/10.1007/s00426-015-0687-4

Verhaeghen, P., & Marcoen, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: evidence for an amplification model. Psychology and Aging, 11(1), 164–178. http://doi.org/10.1037/0882-7974.11.1.164

von Bastian, C. C., & Eschen, A. (2016). Does working memory training have to be adaptive? Psychological Research, 80(2), 181–194. http://doi.org/10.1007/s00426-015-0655-z

von Bastian, C. C., Guye, S., & De Simoni, C. (2017). How strong is the evidence for the effectiveness of working memory training? In M. F. Bunting, J. M. Novick, M. R. Dougherty, & R. W. Engle (Eds.), Cognitive and Working Memory Training: Perspectives from Psychology, Neuroscience, and Human Development. Oxford University Press. Manuscript under review.

von Bastian, C. C., Langer, N., Jäncke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. Memory & Cognition, 41(4), 611–624. http://doi.org/10.3758/s13421-012-0280-7

von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. Behavior Research Methods, 45(1), 108–115. http://doi.org/10.3758/s13428-012-0224-y

von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. Journal of Memory and Language, 69(1), 36–58. http://doi.org/10.1016/j.jml.2013.02.002

von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. Psychological Research, 78(6), 803–820. http://doi.org/10.1007/s00426-013-0524-6

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. Psychonomic Bulletin & Review, 14(5), 779–804. http://doi.org/10.3758/BF03194105

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. Psychonomic Bulletin & Review, 19(6), 1057–1064. http://doi.org/10.3758/s13423-012-0295-x

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? Frontiers in Psychology, 4. http://doi.org/10.3389/fpsyg.2013.00433

Willis, S. L., & Marsiske, M. (1993). Manual for the Everyday Problems Test. University Park: Department of Human Development and Family Studies, Pennsylvania State University.

Zimmermann, K., von Bastian, C. C., Röcke, C., Martin, M., & Eschen, A. (2016). Transfer after process-based object-location memory training in healthy older adults. Psychology and Aging, 31(7), 798–814. http://doi.org/10.1037/pag0000123

Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: effects of age, baseline performance, and training gains. Developmental Psychology, 50(1), 304–315. http://doi.org/10.1037/a0032982